NBER WORKING PAPER SERIES

HOW DOES EXPROPRIATION RISK AFFECT INNOVATION?

Jose-Miguel Benavente
Claudio Bravo-Ortega
Pablo Egaña-delSol
Bronwyn H. Hall

Working Paper 32288
http://www.nber.org/papers/w32288

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
March 2024

We would like to express our gratitude to Gabriel Cruz for his support as a research assistant. PE is grateful for the support provided by the Centre for Social Conflict and Cohesion Studies COES (ANID/FONDAP/15130009). We are very grateful to seminar participants at Universidad Adolfo Ibañez, and the Strategy, Innovation, Entrepreneurship and Sustainability group members for their invaluable comments. We also thank comments from Josh Lerner at Harvard University. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Does Expropriation Risk Affect Innovation?
Jose-Miguel Benavente, Claudio Bravo-Ortega, Pablo Egaña-delSol, and Bronwyn H. Hall
NBER Working Paper No. 32288
March 2024
JEL No. O17,O30,O50

ABSTRACT

We analyze how expropriation risk reduces incentives for innovation and reallocates resources from the innovative sector, building on Romer's(1990) model. Our framework predicts the R&D expenditure, the share of human capital in R&D, the number of patents, technical progress, and economic growth are all lower due to lower expected profits and patent devaluation in the presence of expropriation risks. Empirical analyses, based on a LASSO Instrumental Variable approach and a novel comprehensive dataset spanning nearly two decades, confirm our theoretical predictions. We find robust evidence that expropriation risk, such as corruption, negatively impacts innovation by reducing R&D expenditure, human capital in R&D, number of patents, scientific publications, and the Economic Complexity Index, which is our proxy for technical progress. These findings highlight the detrimental effects of expropriation risk on innovation and economic development at the country level.

Jose-Miguel Benavente
Corporación de Fomento de la Producción
Moneda 921
Santiago
Chile
josemiguel.benavente@corfo.cl

Claudio Bravo-Ortega
School of Business
Universidad Adolfo Ibañez
Diagonal Las Torres 2640
Peñalolén
Santiago
Chile
claudio.bravo@uai.cl

Pablo Egaña-delSol
Escuela de Negocios
Universidad Adolfo Ibanez
Diagonal Las Torres 2640
Peñalolén
Santiago
Chile
pablo.egana@uai.cl

Bronwyn H. Hall
University of California, Berkeley
Economics Department
123 Tamalpais Road
Berkeley, CA 94708
and NBER
bhhall@berkeley.edu

I. Introduction

The concepts of institutions and innovation are central to development economics. The quality of institutions including governmental and innovation have been shown to be a key factor on economic growth and development (North, 1991; Aghion and Howitt, 1992; Murphy, Shleifer, and Vishny, 1993; La Porta et al., 1999; Acemoglu et al., 2001 and 2005; Glaeser et al., 2004, among recent studies). Innovation has large spillovers, which lead to under-investment by the private sector, often leading to government policy intended to encourage innovative activities (Nelson, 1959; Arrow, 1962). For instance, many economists have studied the ways in which subsidies impact the innovative capacity of firms (for recent studies, see Howell, 2017; Wang, Li, and Furman, 2017), or have shown that several breakthrough discoveries during the past century were funded initially by the government (Mowery and Rosenberg (1989).[2]

However, this involvement is a two-edged sword: it helps to reduce certain market-failures (appropriability, access to capital, information frictions) but it also increases the risk of expropriation, such as corruption, clientelism or political favors, by the government, leading to a misallocation of efforts and resources (Cohen and Noll 1991; Fang et al. 2018, Akcigit et al, 2023). Surprisingly, the degree in which expropriation risks or corruption affect innovation and thus economic development have attracted relatively little attention, as a review of some of the papers in this literature suggests (e.g., Bond, Harhoff, and Van Reenen, 2005; Bronzini and Iachini, 2014; Lach, 2002; David, Hall, and Toole 2000). A notable exception is the recent work of Fang et al. (2018), which studies the case of Chinese Research and Development (R&D) subsidies and how an anti-corruption campaign affects their allocation at firm level. In particular, they show that both the anti-corruption campaign and the departures of local government officials responsible for innovation programs strengthened the relationship between firms' historical innovative efficiency and subsequent subsidy awards and depressed the influence of their corruption-related expenditures.

In addition, at the macro-level there are recent empirical studies focusing on the role of institutions (Silve et al., 2018; Rodriguez-Pose et al. 2014; Tebaldi et al. 2013) and the role of financial deepening and democracy (Ho et al., 2018) on innovative outcomes. This literature primarily uses patents as a proxy for innovation. To the best of our knowledge there is no country level analysis on the link between corruption and innovation. Certainly, the relatively scarce research on this topic is noticeable given the importance

[2] See Mazzucato (2013) for interesting examples such as the USA Defense Advanced Project Agency roles in supporting the development of the world's first computer companies and the National Institutes of Health in promoting the fledgling biotech industry.

of innovation for economic growth and the concern that the innovative sector is particularly vulnerable to capture (Murphy, Shleifer, and Vishny, 1993). Reasons for this lack may be the difficulty of constructing a causal analysis because of potential endogeneity between innovation and corruption as well as a lack of appropriate data.

In this article we explore how expropriation risk affects innovative activity, as proxied by R&D expenditures, patent and scientific publication outcomes, human capital allocation, and productive structure, both theoretically and empirically. Our model incorporates the expropriation risk in the innovative sector, while assuming a competitive zero profit sector of final goods sector. We find that this decreases the expected payoff from innovation, lowering the expected profits of innovative companies and the marginal product of human capital devoted to innovation. These results have important implications. First, decreasing the value of expected profits decreases the value of patents in the innovative sector. Second, decreasing the profitability of the sector decreases the value of the marginal product of human capital in the sector, causing some to relocate to the non-innovative sector. This, in turn, decreases R&D expenditure and the technological progress rate and hence the economy's growth. These results hold for the non-innovative sectors whose marginal productivity of human capital is kept fixed or does not decrease as much as in the innovative sector, maintaining the human capital reallocation from the innovative to the non-innovative sector.

Our theoretical model extends Romer's (1990) model by incorporating a generic, linearly multiplicative, "expropriation risk" in the R&D sector. We focus on political control and corruption (i.e. bribes that firms have to pay to bureaucrats or politician) as a particular type of expropriation, but our analysis can be generalized to other types of expropriation risks, such as poor rule of law, arbitrary bureaucratic costs, lack of stable institutional arrangements, low quality of democracy and expropriation itself. Romer's (1990) was the first formal model incorporating the role of R&D as determinant of the human capital allocation on the economy, studying its impact on economic growth. We use this model due to its malleability and formal simplicity that allow us to add the impact of expropriation risks on innovation and to predict theoretically how it could affect human capital allocation in different sectors in the economy, the expenditures in R&D and finally economic growth. We find several interesting predictions: (i) The existence of bribes reduces an innovative firm's profits, and hence the value of innovation patents; (ii) The decrease in profitability decreases the value of the marginal product of human capital in the sector; (iii) This in turn lowers the level of R&D expenditure, the rate of technical progress, and hence economic growth.

Empirically, our analysis uses a novel and comprehensive country-level dataset spanning almost 20 years, from 1996 to 2014, drawn from the following: (i) World Development Indicators (WDI); (ii) International Country Risk Guide (ICRG); (iii) Political Instability Task Force (polity IV), (iv) The Global State of Democracy (GSoD), (v) United Nations Educational, Scientific and Cultural Organization (UNESCO), (vi) World Intellectual Property Organization (WIPO), and (vii) Economic Complexity Index (ECI). Use of these data allow a broad macro-level analysis at the expense of relatively lower suitability for correct identification that micro-level analysis have for identification. Tables 1 to 3 and A1 in the appendix describe the data in terms of time period, geographic zone and source. To correct for measurement error in our corruption variable and to reduce endogeneity, we combine a two stage least squares instrumental variable approach (2SLS), with a machine learning methodology as LASSO (Least Absolute Shrinkage and Selection Operator) to select our set of instruments from those provided by the Polity IV and GSoD datasets. Specifically, we use the LASSO penalized regression in the first stage of the 2SLS model to choose the set of instruments that best predicts the endogenous variables and then use them in the second stage (Belloni et al., 2011; Belloni at al., 2012). Thus, we use variables associated with the political system and state of democracy in the country as instruments for the level of corruption, in an effort to control for the considerable measurement error in that (qualitative) variable as well as potential simultaneity between corruption and innovation. This 2SLS powered by LASSO approach has been recently used in a range of applications to control for endogeneity (e.g., Aral and Nicolaides, 2017; Windmeijer et al., 2019).

Our empirical results are consistent with the main theoretical predictions. Regarding to our Romer (1990)'s based hypotheses, we found evidence consistent with the following results: The existence of corruption is associated with lower domestic patent applications, lower R&D spending, a lower R&D-GDP ratio, lower R&D employment, and lower R&D employment-population share. It does not however reduce the productivity of R&D in terms of patents or scientific papers, possibly because the higher cost of innovation from corruption requires higher average returns. We also found that corruption affects the high technology exports and the Economic Complexity Index (ECI, Hausmann et al., 2013) negatively. The ECI measures the degree of complexity in a country's output matrix in terms of the variety of products and services and we use it along with the high technology export share as proxies for the number of goods varieties produced in the economy, which in Romer's model are related to the rate of technical progress and economic growth.

This article contributes to the literature in two ways. First, we extent Romer's (1990) model in order to add a flexible form of expropriation risk and we derive the optimal conditions and subsequent implications to R&D expenditures, human capital allocation, and economic growth. Second, we contribute to the scarce literature on innovation and corruption by providing country-level empirical evidence in favor of our theoretical implications.

The remainder of the paper is organized as follows: Section II extends the Romer (1990) model incorporating an expropriation risk. Section III describes the data and the empirical strategy in order to test our main hypotheses. Section IV presents the empirical results, while section V concludes.

II. Theoretical Model

Romer's (1990) model is the first formal model that incorporates the role of R&D as determinant of the human capital allocation on the economy, and studies its impact on economic growth. We use this model due to its malleability and formal simplicity that allow us to add in a simple manner the role of bribes or expropriation risks on R&D. Of course, this model is highly stylized, with no depreciation, infinitely lived patents, zero raw labor in the production of intermediate inputs, etc.

In this section, we present our theoretical model, which extends the work of Romer (1990) to the presence of an expropriation risk that we denote by *e*, that can be also interpreted as the bribe that innovative firms pay to corrupt politicians. For ease of exposition we assume that expropriation risk affects the production of new capital goods (intermediate inputs in Romer's model) rather than final demand production. This fee is not a lump sum payment, but is proportional to firms' revenues. As in Romer, the analysis throughout assumes a steady state balanced growth path.

In Romer's model there is an intermediate inputs sector (x) that creates new capital goods through R&D, and a final goods sector (Y). Advanced human capital H moves freely between sectors, with H_Y used in final goods sector and $H_A = H\text{-}H_Y$ available for the production of new designs of intermediate capital goods.

Final output is given by the following modified Cobb-Douglas equation:

$$Y(H_y, L, x) = H_Y^{\alpha} L^{\beta} \int x_i^{1-\alpha-\beta} di \quad (1)$$

This formulation implies that final output is produced by labor, the human capital allocated to the final goods sector, and an aggregate of intermediate inputs that are additively separable in production. The subscript i is a continuous index that counts the number of intermediate inputs varieties, and A is the total number of varieties (the level of non-rival knowledge accumulation). Because of symmetry, all x_i are equal, so we can solve the integral and write the final output equation as:

$$Y = H_Y^{\alpha} L^{\beta} A x^{(1-\alpha-\beta)} \quad (2)$$

Romer (1990) assumes that the intermediate input sector needs η units of final output to produce one unit of a new capital good. This implies that the profit maximizing problem for the innovative firms in the intermediate inputs sector is the following (with r the rental price of capital):

$$Max(1-e)\cdot p_i x_i - r\eta x_i \tag{3}$$

subject to the demand for intermediate input x_i from the final goods sector:

$$p_i = (1-\alpha-\beta)H_Y^{\alpha}L^{\beta}x_i^{-(\alpha+\beta)} \tag{4}$$

Imposing the first order condition for x_i we obtain:

$$(1-e)(1-\alpha-\beta)^2 H_Y^{\alpha}L^{\beta}x_i^{-(\alpha+\beta)} - r\eta = 0 \tag{5}$$

Thus, the optimal intermediate input production amount is reduced by the expropriation e:

$$x *=\cdot \left[\frac{(1-e)(1-\alpha-\beta)^2 H_Y^{\alpha}L^{\beta}}{r\eta}\right]^{\frac{1}{(\alpha+\beta)}} \tag{6}$$

and the monopolistic price is given by:

$$p *= \frac{r\eta}{(1-e)(1-\alpha-\beta)} \tag{7}$$

Replacing both expressions in the firm's profits, this can be expressed as follows:

$$\pi^C = (\alpha+\beta)p_i^* x_i^*(1-e) < \pi^{NC} = (\alpha+\beta)p_i^* x_i^* \tag{8}$$

Where π^{NC} is benchmark profits without corruption. The presence of corruption lowers the profit to each intermediate good producer by e.

Romer (1990) shows that the present value of profits in the intermediate goods sector is simply the profit stream capitalized by the interest rate under the assumptions of his model, which imply a constant present value. Because production and profits under corruption are smaller than without corruption, the present value of profits is also smaller:

$$P_A^C = \frac{\pi^C}{r} = \frac{(\alpha+\beta)p_i^* x_i^*(1-e)}{r} < \frac{\pi^{NC}}{r} = P_A^{NC}$$

Substituting x* into equation (1), we obtain the following equation for final output:

$$Y = H_Y^{\alpha}L^{\beta}A\left[\frac{(1-\alpha-\beta)^2 H_Y^{\alpha}L^{\beta}(1-e)}{r\eta}\right]^{\frac{(1-\alpha-\beta)}{(\alpha+\beta)}} \tag{9}$$

The level of capital K in the economy is a function of the productive capacity of the intermediate inputs:

$$K = A\bar{x}\eta \tag{10}$$

With $\bar{x}$ the production level of the A intermediate inputs, and η a constant factor of conversion between varieties and their productive capital. Note that in the presence of corruption, K will also be lower, because $\bar{x}$ *will be smaller.* At x^*, the profit maximizing choice of $\bar{x}$, we can write the final output equation as a function of K:

$$Y = H_Y^{\alpha} L^{\beta} A^{(\alpha+\beta)} \left(\frac{K}{\eta}\right)^{(1-\alpha-\beta)} \tag{11}$$

Thus, the size of the economy is smaller because capital is lower.

The R&D expense R is the research sector wage times the human capital devoted to R&D:

$$R = w_A H_A \tag{12}$$

$H = H_Y + H_A$ Therefore:

$$Y = \left(H - \frac{R}{w_A}\right)^{\alpha} L^{\beta} A^{(\alpha+\beta)} \left(\frac{K}{\eta}\right)^{(1-\alpha-\beta)} \tag{13}$$

Productivity growth g depends on the production of new goods produced by human capital in the research sector with productivity parameter δ and is given by the following:

$$g = \frac{\dot{A}}{A} = \delta H_A = \frac{\delta R}{w_A} \tag{14}$$

If R&D spending or its productivity is lower, the growth of the economy will be also lower.

We can then show that levels of productivity depend positively on product variety and negatively on expropriation risk, and productivity growth depends positively on R&D expenditure, which in turn is reduced by the expropriation risk by the fact that now there will be fewer workers doing R&D.

To show this we write the equilibrium conditions in the human capital labor market as follows:

$$w_A = w_Y \tag{15}$$

Human capital in the intermediate goods sector receives all the profits in that sector, so the wage is the present value of a new capital good times the quantity of new capital goods produced $P_A \delta A$, which is lower by e. The human capital wage in the final goods sector is its marginal product. Therefore, the equilibrium allocation of human capital is defined by the following:

$$P_A \delta A = P_A^{NC} \delta A(1-e) = \alpha H_Y^{\alpha-1} L^{\beta} A x*^{(1-\alpha-\beta)} \tag{16}$$

Thus, given that with expropriation risk the value of the marginal productivity of workers in research is smaller, there will be more workers in the production of final goods and fewer in research. In fact H_Y is increased by the inverse of *(1-e):*

$$H_Y = \frac{\alpha r}{(1-\alpha-\beta)(\alpha+\beta)\delta(1-e)} = \frac{\Lambda r}{\delta(1-e)} \quad (17)$$

Equation (17) implicitly defines *Λ = α/[(1-α-β)(α+β)].* Productivity growth is given by

$$g = \delta H_A = \delta H - \frac{\Lambda r}{(1-e)} \quad (18)$$

Thus the higher the expropriation risk the lower the amount of human capital doing R&D. The R&D expenditure can be computed as follows:

$$R = w_A \left(H - \frac{\Lambda r}{\delta(1-e)}\right) = P_A^{NC} \delta A (1-e) \left(H - \frac{\Lambda r}{\delta(1-e)}\right) \quad (19)$$

Hence, R&D levels are lower.

If the model is closed with a utility function for the representative consumer, we can determine the steady state interest rate, which is now also lower:

$$r = \frac{(\delta H + \rho/\theta)\theta(1-e)}{1-e+\Lambda} \quad (20)$$

With ρ the representative consumer discount factor and θ the consumption elasticity in a CES utility function.[3]

Therefore, the growth rate is also lower:

$$g = \frac{1}{\theta}\left(\frac{\delta H(1-e) - \rho\Lambda}{1-e+\Lambda}\right) \quad (21)$$

The conclusions of this model are that 1) the existence of bribes reduces an innovative firm's profits, and hence the present value of its innovations. 2) The decrease in innovation values decreases the value of the marginal productivity of human capital in R&D. 3) The lower innovation value decreases the allocation level of human capital to R&D in steady state, decreasing therefore the level of R&D expenditure, the rate of technical progress, interest rates and hence economic growth.

III. Data

In the empirical section, we explore the effect that corruption may have on innovative activity using a large panel of countries between 1996 and 2014. In this section we briefly present our database and we show the link between corruption and innovation efforts. A more detailed description of the data is given in the Appendix.

---

[3] Where $u(c) = \frac{c^{1-\theta}-1}{1-\theta}$

III.A Key variable: Expropriation Risk

As suggested by Fang et al. (2018) corruption is hard to measure. There are only few sources of information that could compare different countries in terms of levels of corruption. Not only because information is not fully available but, as suggested previously, corruption has several dimensions and it hard to see which is more related with the risk of government expropriation.

One standard source of comparable figures is given by the International Country Risk Guide (ICRG). This rating compares near twenty different dimensions to provide a means of assessing the financial, economic and political stability of countries on a comparable basis. In particular, corruption is a subcomponent of the political stability index. For the ICRG, this component is presumed to act as a threat to foreign investment for a number of reasons:, such as demands for special payments and bribes connected with import and export licenses, exchange controls, tax assessments, police protection, or loans. The ICRG emphasizes the risks arising from distortion of the economic environment due to enabling people to assume positions of power through patronage (nepotism, favor-for-favor, secret party funding and close politics-business ties) instead of ability that may introduce inherent instability into the political process.

The corruption index is created by evaluating a country's practices and conducts by asking if there are demands for special payments and bribes, exchange controls, tax assessments or police protection for those involved with trade and financial transactions. In addition, other dimensions like patronage, nepotism, job reservations, secret party funding, and ties between politics and business are also considered.[4]

ICRG standardizes the variable between 0 (most corrupt) and 6 (least corrupt) in order to make it comparable among countries and also between years; in our analysis we reverse the scale so that 6 is the highest level of corruption and 0 the lowest, which makes the regressions somewhat easier to interpret. In Table 1 we present simple statistics for this variable for the whole sample of countries for which there is available information. The lower panel of Table 1 presents statistics for the sample with R&D data available and that we use in the subsequent analysis. This subsample has slightly lower values of the index (i.e. less corruption). The variance of the index is approximately the same across countries and within countries, which suggests that there will be identifying power even with country fixed effects, as corruption varies across countries over time. The index shows no strong trend, however, with the exception of a slightly lower median in the first couple of years (not shown).

[4] For more details see https://www.prsgroup.com/explore-our-products/international-country-risk-guide/

[Table 1 about here]

The histogram in Figure 1 shows that the index takes on a discrete number of values, slightly skewed toward 6, so it is a fairly coarse measure. In what follows, we explore the use of instrumental variables to improve the predictive power of this measure.

[Figure 1 about here]

III.B Instrumental variables

As discussed above, corruption and the related expropriation risk are difficult to measure. The related literature (See for example the recent review of Mungiu-Pippidi & Heywood, 2020; Khwaja et al. 2015; Akcigit et al., 2023) suggests that there are several indirect ways to capture the level of corruption at country level. In what follows, we will explore the use of instrumental variables from other sources that are plausibly related to corruption.

We select two different sources that have been used in the prior literature and that have a large number of countries with data available. First, we use the Polity Project (Marshall et al. 2019), which contains data on the form of government (democracy or autocracy), the process of choosing the chief executive, and access or participation in the political system. Second, we use the Global State of Democracy dataset (Tufis 2018), which contains a broader set of data that includes some media openness variables, and information on access to the courts and justice. In Table 2 we present these variables and their sources. All variables have been recoded so that the expected correlation with the index of corruption is positive. For example, this means that a variable like *civil_participation* indicates a lack of civil participation.

[Table 2 about here]

In the data appendix Table A3 we present the correlations between expropriation risk and these proxies of corruption. In general, most of the proxies are highly correlated with the corruption index and each other. This is perhaps not surprising given the procedures and sources used to construct these variables. Our preliminary explorations suggested that use of all these variables as instruments along with fixed country effects led to imprecise estimates given this correlation and the qualitative nature of the variables. Therefore, we chose to use LASSO (Least Absolute Shrinkage and Selection Operator) estimation methods when including these proxy variables as instruments for corruption.

III.C Dependent Variables: scientific efforts and innovation results

Following Hall and Rosenberg (2010) and Qureshi et al. (2020), there are several and complementary ways to measure innovation efforts at a country level. Most of them are related to financial and knowledge inputs like R&D expenditure and the number of R&D researchers, and few others related to intermediate outputs like scientific papers and new patents. In order to capture the (potentially variable) effects of expropriation risk on the whole innovation process we will use both input and intermediate output variables. In what follows we list the included variable and the source of information from which the variable was obtained or constructed.

R&D expenditure information is obtained from the World Development Indicator database; it is also available from UNESCO.[5] We use two variables for expenditure: the expenditure itself (GERD, measured in PPP dollars) and the expenditure as a share of GDP. The number of full time equivalent researchers engaged in Research & Development (R&D) as a share of population in millions is also obtained from WDI/UNESCO data.[6] Researchers are defined as professionals who conduct research and improve or develop concepts, theories, models techniques instrumentation, software of operational methods in basic research, applied research, and experimental development.

The data on scientific and technical journal articles refer to the number of scientific articles published by origin country. These data are obtained from the World Development Indicators from the World Bank, although their ultimate source is data collected by the U. S. National Science Foundation.[7] These data are only available from the year 2000 forward.

Patent applications statistics are obtained from the World Intellectual Property System (WIPO) database.[8] WIPO provides data on applications to each country's IP or patent office by residents of the country and non-residents separately. We use the resident applications, as these are closest to the concept of inventions produced in the country.

We utilize two additional variables to gauge the success of innovative output. The Economic Complexity Index (ECI), introduced by Hausmann et al. (2013), serves as a comprehensive metric of the productive capabilities of large economic entities such as cities, regions, or countries. Specifically, the ECI aims to elucidate the collective knowledge amassed within a population and manifested through the economic activities within a given area. To achieve this objective, the ECI defines the knowledge within a location as the average knowledge level of the activities it hosts, and the knowledge associated with a particular

---

[5] United Nations Educational, Scientific and Cultural Organization (UNESCO) data is available at http://data.uis.unesco.org/

[6] Both researchers per population and R&D intensity are also available from the World Development Indicator database, but their source is also UNESCO.

[7] The World Development Indicators (WDI) from the World Bank is available at https://datatopics.worldbank.org/world-development-indicators/

[8] WIPO statistics are available at https://www3.wipo.int/ipstats/

activity as the average knowledge level of the locations where that economic activity occurs. Mathematically, the ECI is rigorously defined in terms of an eigenvalue derived from a matrix connecting countries to countries, which is a projection of the matrix linking countries to the products they export. By incorporating information on both the diversity of countries and the ubiquity of products, the ECI offers a measure of economic complexity that encompasses insights into both the diversity and sophistication of a country's exports.[9] We also use the IMF measure of export diversity sourced from The Diversification Toolkit: Export Diversification and Quality Databases (IMF, 2018).[10]

In Table 3 we present simple statistics for all the variables that characterize the scientific effort and innovation from those countries for which we have comparable information. Table 4 presents some details for these same variables in logarithmic form. In particular we show the variance decomposition between and within countries. As can be observed, most of the variation is explained between countries rather than within each country throughout the period we are considering.

[Tables 3 and 4 about here]

In order to take a first look at the relationship between innovative activity and corruption, in Figure 2 we present the simple correlation between our corruption index and some innovative activity indicators. The standard error bands on the graphs below are based in standard errors clustered on country.

[Figure 2 about here]

As can be observed there is a strong negative correlation between most of the selected variables and our variable of interest. Relative to GDP and population R&D expenditure and the number of R&D researchers differ by 100 per cent between the lowest and highest values of the corruption index. The economic complexity index varies from -0.5 to 0.5 over the range and the share of exports in the high tech sectors by 120 per cent. Interestingly, the production of patents by residents does not fall given the level of R&D, although of course it is much lower when R&D is not controlled for; this variable moves in parallel with R&D, suggesting a limited separate effect from corruption on innovative output. To a lesser extent, the same is true for scientific and technical article production relative to R&D, which increases only by 10 per cent over the range of the corruption index. Table A5 in the appendix shows that even this small positive effect disappears when we control for year and country effects.

Our final samples for estimation vary in coverage, due to differences in data availability for the various innovation variables. Table A4 in the appendix lists the 130 countries for which we had more than one

[9] Currently available at The Observatory of Economic Complexity (oec.world)

[10] Available at https://www.imf.org/external/np/res/dfidimf/diversification.htm

year of data on corruption and the quality of governance variables. Of these, 10 had no data on any of the innovation indicators except the economic complexity index. 106 had at least some data on R&D spending, while only 88 had data on the number of researchers. 107 had data on the high technology export share. When estimating, we used four different samples corresponding to the columns of Table A4, on the grounds that we wanted as large a coverage as possible.

IV. Results

We now move to test our main hypothesis derived from our theoretical model. Our goal is to test if corruption has an effect on scientific effort and/or innovation output at a country level. Following the standard literature we estimate the following equation:

$$y_{it} = \beta x_{it} + \nu_i + \lambda_t + \varepsilon_{it} \tag{22}$$

where y_{it} captures the different scientific or innovation variables for country i at year t; x_{it} is our corruption variable that also varies among countries and through time, ν_i captures country fixed effects, and λ_t controls for the overall time fixed effects. Our estimated standard errors were robust to heteroskedasticity and serial correlation within countries.

Ordinary least squares results, both with and without country fixed effects are presented in Appendix Table B1. As also shown in Figure 2, the cross-country results without fixed effects show a large negative relationship of corruption to the levels of R&D, patents, scientific articles, high technology export share, and economic complexity, albeit with good-sized standard errors. When patents and scientific articles are scaled by R&D investment, the negative impact of corruption disappears. Overall, these results imply an approximately 20 per cent decline in the various innovation rates with each one unit change in the corruption index. Including country effects both reduces the magnitude of the coefficients and makes them insignificant, with much smaller standard errors. The conclusion is that the main impact of corruption levels is a permanently lower innovation rate in a country. However, given the potential for many omitted country characteristics in this relationship, we have preferred to focus on estimates with country fixed effects in what follows.

As we indicated earlier, the corruption index is a qualitative measure and is fairly coarse, suggesting substantial measurement error in our key right hand side variable. Therefore, in the next section we present results using a larger list of governance and political system variables to instrument the corruption index.

IV.A Two Stage Least Squares (2SLS) estimation model

The OLS results in the appendix suggest that measurement error in the corruption index (which is admittedly fairly subjectively determined) may be a source of bias in our coefficient estimates. To explore this possibility, we collected a large set of instruments that are plausibly related to expropriation risk, as described in section III.B.

The results of two-stage least squares estimation applied to equation (22) are shown in Appendix Tables B2 and B3. Table B2 shows the first stage results, both with and without country effects. Without country effects and due to multicollinearity, few variables are statistically significant, but the R-squared is a reasonably high 0.119, given the qualitative nature of the variables. When we include country effects, the within R-squared is 0.053. Within countries, the most important variables predicting corruption are a lack of participation in civil society and lack of access to a fair legal system.

Table B3 shows the second stage estimates, with and without country effects. Without country effects, corruption has a very large negative impact on almost all the innovation variables, with large standard errors. Although the test for overidentification passes easily, the test for weak instruments fails, except in the cases of high tech export shares and the economic complexity index. Even there, the test would fail if the critical values from Stock and Yogo (2005) were used. The situation is only slightly better when country effects are included. The bottom panel of Table B3 shows that corruption is now a small and insignificant predictor of all the innovation variables when adding country fixed effects. The test of overidentification still passes and now the weak identification test passes marginally for the R&D spending and patent variables. Our conclusion was to search for an estimation method more robust to collinearity of the instruments in a finite sample.

When there are a large number of instrumental variables, Belloni et al (2011) and Belloni et al (2012) suggested it is useful to use a LASSO (Least Absolute Shrinkage and Selection Operator) estimator to select the instruments from a pool of candidates. The argument is that because the goal of the first stage in 2SLS is to predict the endogenous variable in the second stage, in finite samples it is useful to choose method like LASSO that is good at prediction while minimizing the possibility of overfitting.

In the context of a panel structure, the Two Stage Least Squares is defined by:

$$X_{it} = \alpha_0 + \alpha_1 Z_{it}^1 + \alpha_2 Z_{it}^2 + \upsilon_{it} \quad (23)$$

$$y_{it} = \beta_0 + \beta_1 \hat{X}_{it} + \varepsilon_{it} \quad (24)$$

where X_{it} is the corruption variable, Z_{it1} are the instrumental variables associated with corruption, and y_{it} is the innovation variable.

Then, following Belloni et al. (2012) we use a LASSO penalized regression in the first stage of the 2SLS model to choose the set of instruments that best predicts the endogenous variable and then use them in the two stage least squares model. In particular, the LASSO (Tibshirani, 1996), minimizes the sum of squares of the residuals plus a penalty function for the variance of the coefficient estimates that shrinks the estimates towards zero.

There is a cost in terms of precision to including many instruments, and the LASSO method reduces the objective function by throwing out the instruments that contribute little to the fit. The effect of the penalization is that LASSO sets the coefficients for some variables to zero.

IV.B LASSO Results

Results of the LASSO model are presented in Table 5. A large number of control variables are excluded by the procedures. First, we note that none of the year dummies enter any of the equations, which implies that corruption controlling for governance quality has not changed over time. The raw corruption data showed only a slight trend (not shown). Second in each equation, only slightly less than half of the country dummies were retained. In general, the retained dummies were those for larger economies such as the USA, China, Germany, the UK, and Japan. Dropping the dummies for the large number of smaller countries produced the result that the LASSO IV estimates were roughly the same with and without the country effects, although the standard errors are somewhat larger when the limited number of country dummies is included.

With respect to the excluded instruments, usually only 4 or fewer of them were retained: access to justice, civil participation, harassment of journalists, and whether chief executive recruitment is regulated and standardized in the country. The only exception was the ECI dependent variable, where the competitiveness of participation in politics and the level of democracy were added to the previously listed instrument set. For the two scientific and technical publication variables, only the harassment of journalists survived as an excluded instrument. These results agree roughly with those for the first stage in Table A6.

The results themselves resemble those for IV without country effects (top panel of Table A7). In most of the regressions, corruption has a negative effect on the various innovation indicators. Consistent with our theoretical predictions, we find that corruption negatively affects innovation by reducing R&D expenditure, human capital in R&D, number of patents, scientific publications, and the high tech exports and the Economic Complexity Index. There are two exceptions to this pattern, the yield of patent applications and of scientific and technical articles from R&D. Both are slightly positive and marginally significant. In other words, corruption may reduce innovative effort, but not its productivity. In fact, to the extent that more productive or profitable projects are chosen, it may slightly increase innovation productivity on the margin.

[Table 5 about here]

V. Conclusions

This paper sheds light on the intricate relationship between institutions, innovation, and economic development. The nexus between the quality of institutions, particularly governments, and innovation has been extensively explored in the literature, underscoring their pivotal role in fostering economic growth (North, 1991; Aghion and Howitt, 1992; Murphy, Shleifer, and Vishny, 1993; La Porta et al., 1999; Acemoglu et al., 2001 and 2005; Glaeser et al., 2004). However, the involvement of governments in promoting innovation presents a double-edged sword, as it addresses market failures but also introduces risks of expropriation, such as corruption and political favors (Cohen and Noll, 1991; Fang et al., 2018).

While the literature has made significant strides in understanding the impact of institutions and financial factors on innovation, there remains a noticeable gap concerning the effect of expropriation risks, particularly corruption, on innovative activity (Bond, Harhoff, and Van Reenen, 2005; Bronzini and Iachini, 2014; Lach, 2002; David, Hall, and Toole, 2000). This paper studies both theoretically and empirically how expropriation risks, manifested through corruption and political control, influence R&D expenditures, patent outcomes, human capital allocation, and economic growth.

Theoretically, we extend Romer (1990)'s model to incorporate expropriation risks in the R&D sector, elucidating the mechanisms through which corruption diminishes the incentives for innovation and reallocates resources away from the innovative sector. Our model predicts a decrease in R&D expenditure, human capital in R&D, number of patents, technical progress, and economic growth due to lower expected profits and the devaluation of patents in the presence of expropriation risks.

Empirically, our analysis utilizes a comprehensive dataset spanning nearly two decades and employs advanced econometric techniques to address endogeneity and measurement error in corruption variables. Consistent with our theoretical predictions, we find that corruption negatively affects innovation by reducing R&D expenditure, human capital in R&D, number of patents, scientific publications, and the high-tech exports and Economic Complexity Index. These empirical findings provide robust support for the hypothesized link between expropriation risks and innovative activity at the country level.

In summary, this article contributes to the existing literature by offering theoretical insights and empirical evidence on the detrimental impact of corruption on innovation and economic development. By highlighting the mechanisms through which expropriation risks distort incentives and resource allocation in the innovative sector, this research underscores the potential importance of addressing governance challenges to foster a conducive environment for innovation and sustainable growth. Moving forward, further research could explore additional dimensions of expropriation risks and their implications for innovation policy and economic performance.

VI. References

Acemoglu, Daron, Simon Johnson, and James A. Robinson (2001). The Colonial Origins of Comparative Development: An Empirical investigation, *American Economic Review*, 91(5): 1367-1401.

Acemoglu, Daron, Simon Johnson ,and James A. Robinson (2005). Institutions as the Fundamental Cause of Long-Run Growth. P. Aghion and S. N. Durlauf (eds.), *Handbook of Economic Growth* Volume 1, Part A, 385-472.

Akcigit, Ufuk, Salome Baslandze, and Francesca Lotti (2023), Connecting to Power: Political Connections, Innovation, and Firm Dynamics. *Econometrica 9*1: 529-564. https://doi.org/10.3982/ECTA18338

Aghion, Philippe, and Stephen N. Durlauf (eds.) (2005). *Handbook of Economic Growth*. New York and Amsterdam: Elsevier.

Aghion, Philippe and Peter Howitt (1992). A Model of Growth through Creative Destruction, *Econometrica,* 60(2): 323-351.

Aral and Nicolaides (2017). Exercise contagion in a global social network. *Nature Communications*, 2017.

Arrow, Kenneth (1962). Economic Welfare and the Allocation of Resources for Invention. In Richard R. Nelson (ed.), *The Rate and Direction of Inventive Activit*y. Princeton: Princeton University Press, 609-625.

Barro, R. J., and Sala-i-Martin, X. (1995). *Economic growth*. Cambridge, MA: MIT Press.

Alexandre Belloni, Daniel Chen, Victor Chernozhukov, Christian Hansen

Belloni, Alexandre, Daniel Chen, Victor Chernozhukov, and Christian Hansen (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica,* 80(6), 2369-2429.

Belloni, Alexandre, Victor Chernozhukov, and Lie Wang (2011). Square-root lasso: pivotal recovery of sparse signals via conic programming. *Biometrika*, 98(4), 791-806.

Bond, Stephen, Dietmar Harhoff, and John Van Reenen (2005). Investment, R&D and Financial Constraints in Britain and Germany, *Annales d'Economie et de Statistique* 79-80, 433-60.

Bronzini, Raffaello, and Eleonora Iachini (2014). Are Incentives for R&D Effective? Evidence from a Regression Discontinuity Approach, *American Economic Journal: Economic Policy* 6, 100-134.

CID, Harvard (2018). *Atlas of Economic Complexity*. https://atlas.cid.harvard.edu/

Cohen, Linda R., and Roger G. Noll (1991). *The Technology Pork Barrel.* Washington: Brookings Institution.

David, Paul A., Bronwyn H. Hall, and Andrew A. Toole (2000). Is public R&D a complement or substitute for private R&D? A review of the econometric evidence. *Research Policy* 29(4): 497-529.

Fang, Lily, Josh Lerner, Chaopeng Wu, and Qi Zhang (2023). Anticorruption, Government Subsidies, and Innovation: Evidence from China. *Management Science* 69(8):4363-4388..

Glaeser, Edward L., Rafael La Porta, Florencio Lopes-de-Silanes, and Andrei Shleifer (2004). Do Institutions Cause Growth?, *Journal of Economic Growth*, 9(1), 271-303.

Hall, Bronwyn H., and Rosenberg, Nathan (2010). *Handbook of the Economics of Innovation*. New York and Amsterdam: Elsevier. ISBN: 9780444519955

Hausmann, Ricardo, César A. Hidalgo, Sebastián Bustos, Michele Coscia, Alexander Simoes, and Muhammed A. Yıldırım (2013). *The Atlas of Economic Complexity*. Cambridge, MA: MIT Press.

Ho, Chun-Yu, Shaoqing Huang, Hao Shi, and Jun Wu (2018). Financial deepening and innovation: The role of political institutions. *World Development*, 109: 1-13.

Howell, Llewellyn D. (2020). *ICRG Methodology*. Liverpool, NY: PRS Group. Available at https://www.prsgroup.com/explore-our-products/international-country-risk-guide/

Howell, Sabrina T. (2017). Financing Innovation: Evidence from R&D Grants, *American Economic Review* 107: 1136-1164.

ICRG (International Country Risk Guide) (2021). Available at https://www.prsgroup.com/explore-our-products/international-country-risk-guide/

IMF (2018). *The Diversification Toolkit: Export Diversification and Quality Databases*, International Monetary Fund. Available at https://www.imf.org/external/np/res/dfidimf/diversification.ht

Khwaja, Asim Ijaz, and Atif Mian (2005). Do Lenders Favor Politically Connected Firms? Rent Provision in an Emerging Financial Market. *Quarterly Journal of Economics* 120: 1371—1411.

La Porta, Rafael, Florencio Lopes-de-Silanes, Andrei Shleifer and Robert Vishny (1999). The Quality of Government. *Journal of Law, Economics, and Organization*, 15(1): 222-279.

Lach, Saul (2002). Do R&D Subsidies Stimulate or Displace Private R&D? Evidence from Israel. *Journal of Industrial Economics* 50, 369-90

Lerner, Josh (1999). The Government as Venture Capitalist: The Long-Run Effects of the SBIR Program, *Journal of Business* 72, 285-318.

Marshall, Monty G., Ted Robert Gurr, and Keith Jaggers (2019). *Dataset Users' Manual. Polity IV Project: Political Regime Characteristics and Transitions, 1800-2018*. Vienna, VA: Center for Systemic Peace. Available at http://www.systemicpeace.org/inscrdata.html

Mowery, David C. and Nathan Rosenberg (1989). Postwar Federal Investment in Research and Development. In Mowery and Rosenberg, *Technology and the Pursuit of Economic Growth*. Cambridge, UK: Cambridge University Press, 123-168.

Mungiu-Pippidi, Alina, and Paul M. Heywood (eds.) (2020). *A Research Agenda for Studies of Corruption*. Cheltenham, UK: Edward Elgar Publishing. Retrieved May 13, 2022, from https://www.elgaronline.com/view/edcoll/9781789904994/9781789904994

Murphy, Kevin M., Andrei Shleifer, and Robert W. Vishny (1993). Why is rent-seeking so costly to growth?. The *American Economic Review*, 83(2): 409-414.

Nelson, Richard R. (1959). The Simple Economics of Basic Scientific Research. *Journal of Political Economy* 67, 297-306.

North, Douglas C. (1991). Institutions. *Journal of Economic Perspectives* 5(1): 97-112.

Qureshi, Irfan, Donghyun Park, Gustavo Atilio Crespi, and Jose Miguel Benavente (2020). Trends and Determinants of Innovation in Asia and the Pacific and Latin America and the Caribbean?. Manila, Philippines: Economic Research and Regional Cooperation Department (ERCD), Asian Development Bank.

Rodriguez-Pose, Andres, and Marco Di Cataldo (2015). Quality of government and innovative performance in the regions of Europe. *Journal of Economic Geography* 15(4): 673-706.

Romer, Paul (1990). Endogenous Technological Change. *Journal of Political Economy* 98: S71-S102.

Silve, Florent, and Alexander Plekhanov (2018). Institutions, innovation and growth: Evidence from industry data. *Economics of Transition* 26(3): 335-362.

Stock, James H., and Motohiro Yogo (2005). Testing for weak instruments in linear IV regression. In D. W. K. Andrews and J. H. Stock (eds.), *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, 80-108. Cambridge: Cambridge University Press.

Tebaldi, Edinaldo, and Bruce Elmslie (2013). Does institutional quality impact innovation? Evidence from cross-country patent grant data. *Applied Economics* 45(7): 887-900.

Tufis, Claudiu D. (2018). *The Global State of Democracy Indices Codebook*. Stockholm, Sweden: International Institute for Democracy and Electoral Assistance (International IDEA).

UNESCO (United Nations Educational, Scientific and Cultural Organization) Institute for Statistics (2022). Database available at http://data.uis.unesco.org/

Wallsten, Scott J. (2000). The Effects of Government-Industry R&D Programs on Private R&D: The Case of the Small Business Innovation Research Program. *RAND Journal of Economics* 31: 82-100.

Wang, Yanbo, Jizhen Li, and Jeffrey L. Furman (2017). Firm Performance and State Innovation Funding: Evidence from China's Innofund Program. *Research Policy* 46: 1142- 1161.

Windmeijer, Frank, Helmut Farbmacher, Neil Davies, and George Davey Smith (2019). On the use of the lasso for instrumental variables estimation with some invalid instruments. *Journal of the American Statistical Association* 114(527): 1339-1350.

WIPO (World Intellectual Property Organization) IP Statistics Data Center (2022). Available at https://www3.wipo.int/ipstats/

World Bank (2018). *World Development Indicators*. https://datatopics.worldbank.org/world-development-indicators/

Table 1

Corruption variable - panel data

Variable		*Mean*	*Std. Dev.*	*Min*	*Max*
Corruption - all observations	overall	3.07	1.31	0	6
	between		0.79	0.58	5.16
	within		1.05	-1.00	8.00
	2660 observations on 140 countries, 1996-2014				
Corruption for sample with R%D info	overall	2.90	1.32	0	6
	between		0.86	0.50	5.00
	within		1.04	-0.86	7.83
	1,448 observations on 108 countries, 1996-2014				

Source: ICRG, PRS Group; authors' computations

Table 2
Corruption Variables

Variable	*Source*	*Description*	*Obs*	*Mean*	*Std. Dev.*	*Median*	*Min*	*Max*
corruption	ICRG	Index based on political risk factors (std)	2660	-0.045	0.954	-0.099	-2.281	2.082
autoc	Polity#	Institutionalized autocracy	2478	-0.066	0.967	-0.682	-0.682	2.735
comp_exec_p	Polity	Competitiveness of chief exec recruitment	2478	-0.108	0.964	-0.899	-0.899	1.904
democ	Polity	Institutionalized democracy	2478	-0.095	0.987	-0.394	-1.168	1.411
open_exec	Polity	Openness of chief executive recruitment	2478	-0.070	0.929	-0.463	-0.463	2.455
parcomp	Polity	Competitiveness of participation in govt	2478	-0.065	0.974	-0.391	-1.929	1.915
parreg	Polity	Regulation of participation in govt.	2478	-0.041	1.035	0.226	-1.480	1.932
reg_exec	Polity	Regulation of chief executive recruitment	2478	-0.109	0.949	-0.848	-0.848	2.631
access_Justice	GsoD*	Access to justice	2489	0.394	0.184	0.414	0	0.886
barriers_parties	GsoD	Barriers to political parties	2508	0.410	0.270	0.368	0	1.000
clean_election	GsoD	Freedom of elections from irregularities	2508	0.416	0.264	0.379	0	1.000
check_govern	GsoD	Checks on government	2508	0.431	0.203	0.408	0	0.944
civil_Liberties	GsoD	Civil liberties	2508	0.322	0.204	0.293	0	1.000
civil_part	GsoD	Civil society participation	2508	0.389	0.193	0.366	0	0.999
comp_exec_g	GsoD	Competitiveness of chief exec recruitment	2432	0.385	0.437	0.000	0	1.000
comp_part	GsoD	Competitiveness of alternative policies	2431	0.306	0.263	0.200	0	1.000
court_indep	GsoD	High court independence	2502	0.446	0.295	0.421	0	1.000
critical_print	GsoD	Critical print, broadcast media	2508	0.368	0.204	0.343	0	1.000
harass_journ	GsoD	Harassment of journalists	2508	0.432	0.195	0.440	0	1.000

Source: The Policty IV Project, Center for Systemic Peace, Marshall et al. (2019). The corruption and polity variables have been standardized. Higher values are associated with more corruption.

* Source: Tufis (2018): Global State of Democracy Indices. Data for 128 countries, 1996-2014, higher values associated with more corruption.

A few observations are missing from GSoD during periods of political upheaval and war, mostly in Africa.

Table 3

Statistics for the innovation variables

Variable	*Source*	*# obs.*	*Mean*	*Std.dev.*	*Median*	*Min*	*Max*
R&D expenditure (PPP$2015M)	WDI	1448	13,314	49,519	738	0.1	482,423
Number of R&D researchers	WDI	1143	91,597	218,454	18,985	58.6	1,558,403
Resident patent apps	WIPO	2383	8,632	45,941	67	0.0	801,135
Sci. & tech. journal articles*	WDI	1885	12,817	43,274	712	0.0	433,192
R&D -GDP ratio (per cent)	WDI	1448	0.96	0.93	0.62	0.005	4.4
Researchers per pop. In millions	WDI	1143	1,891.7	1,782.7	1,428.3	7.3	7,821.9
Resid. patent apps per M R&D	WIPO	1448	1.19	2.38	0.43	0.0	31.3
Sci. & tech. articles per M R&D*	WDI	1202	5.15	13.23	2.93	0.00	286.7
High tech export share (per cent)	IMF	1887	0.05	0.08	0.02	0.0	0.6
Economic complexity index	ECI	2394	0.04	1.04	-0.06	-3.9	2.9
Diversity	IMF	2360	3.24	1.25	3.03	1.23	6.4

108 countries, 1996-2014, unbalanced panel.

Basic sample is countries with good corruption, polity, and GSoD data.

* The number of scientific and technical articles are only available beginning in 2000.

Table 4

Panel satistics for R&D/innovation variables

Variable	*Mean*	*Std. dev. total*	*between*	*within*	*Var ratio within/total*	*Number observations*	*Number countries*	*Average Years*
Log R&D*	7.085	2.415	2.469	0.422	0.031	1422	105	13.5
Log R&D researchers	9.900	2.009	2.132	0.329	0.027	959	85	11.3
Log resident patents*	5.564	3.236	3.26	0.994	0.094	1424	102	14.0
Log sci tech articles	7.896	2.340	2.487	0.398	0.029	1044	97	10.8
Log R&D/GDP*	-0.566	1.158	1.181	0.316	0.074	1441	106	13.6
Log researchers/pop	7.004	1.463	1.740	0.300	0.042	959	85	11.3
Log res. pats per R&D*	-1.522	1.874	1.752	0.986	0.277	1405	101	13.9
Log aritcles per R&D	0.719	0.735	0.776	0.358	0.237	1044	97	10.8
Log hitech export share	0.864	1.874	2.212	0.617	0.108	1317	95	13.9
Log diversity/complexity	4.990	0.557	0.633	0.097	0.030	701	62	11.3

*R&D in PPP is unavailable for Cuba due to lack of exchange rate information. This means we can compute the R&D-GDP ratio and the number of patents for Cuba, but we cannot compute R&D in PPP terms, or the number of patents per R&D in PPP terms, accounting for the difference in observation and country counts for these variables.

Table 5

Results using Cross-fit Instrumental Variable Regression

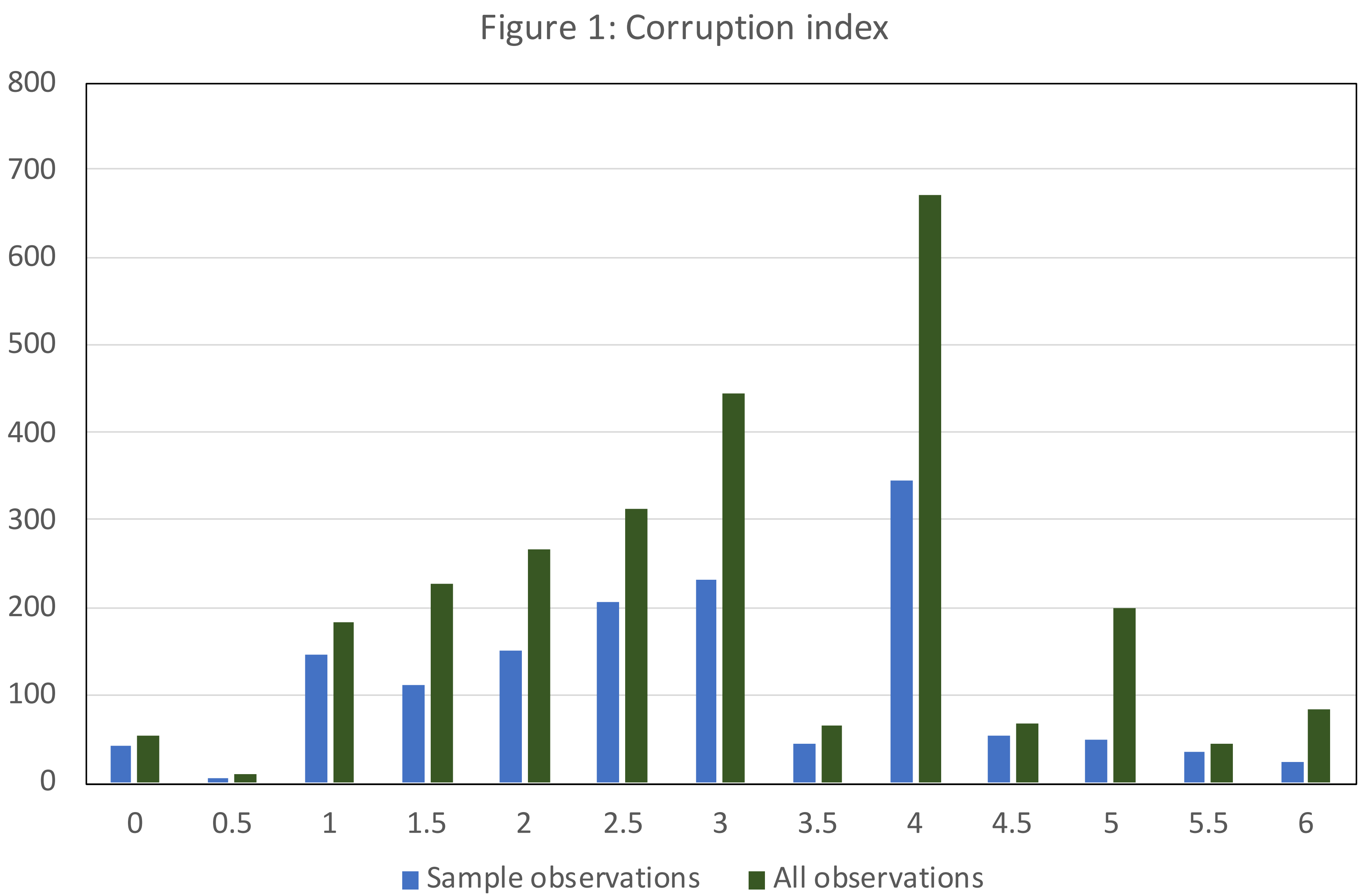
VARIABLES	(1) Log R&D current PPP$	(2) Log R&D-GDP ratio	(3) Log resident patent apps	(4) Log res. patent apps per R&D	(5) Log sci & tech journal articles	(6) Log scitech articles per R&D	(7) Log number of researchers	(8) Log researchers per population	(9) High tech export share	(10) Economic complexity index
					Pooled LASSO IV					
Standardized corruption index	-2.014***	-2.410***	-4.069***	0.540***	-5.448***	1.320***	-2.786***	-3.902***	-3.884***	-2.348***
	(0.248)	(0.254)	(0.477)	(0.190)	(0.958)	(0.275)	(0.455)	(0.524)	(0.416)	(0.188)
Observations	1,446	1,446	1,446	1,446	1,199	1,199	1,134	1,134	1,882	2,391
Number of years selected#	1	0	0	0	0	0	0	0	0	0
Number of instruments selected@	3	4	4	4	2	2	4	4	5	6
Wald Test Prob > chi2	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000
					LASSO IV with country effects					
Standardized corruption index	-3.959***	-2.431***	-3.685***	0.217	-6.852*	1.355*	-3.090***	-4.073***	-2.191***	-2.238***
	(0.594)	(0.351)	(0.607)	(0.187)	(3.610)	(0.692)	(0.778)	(0.893)	(0.320)	(0.212)
Observations	1446	1446	1446	1446	1199	1199	1134	1134	1882	2391
Number of controls selected#	43	45	39	39	21	22	30	34	60	70
Number of instruments selected@	4	4	4	4	1	1	4	4	3	6
Wald Test Prob > chi2	0.000	0.000	0.000	0.244	0.058	0.051	0.000	0.000	0.000	0.000

Standard errors robust to heteroskedacticity and autocorrelation in parentheses

*** p<0.01, ** p<0.05, * p<0.1

#Potential controls are 18 year dummies in the first panel and 18 year and up to 130 country dummies in the second panel.

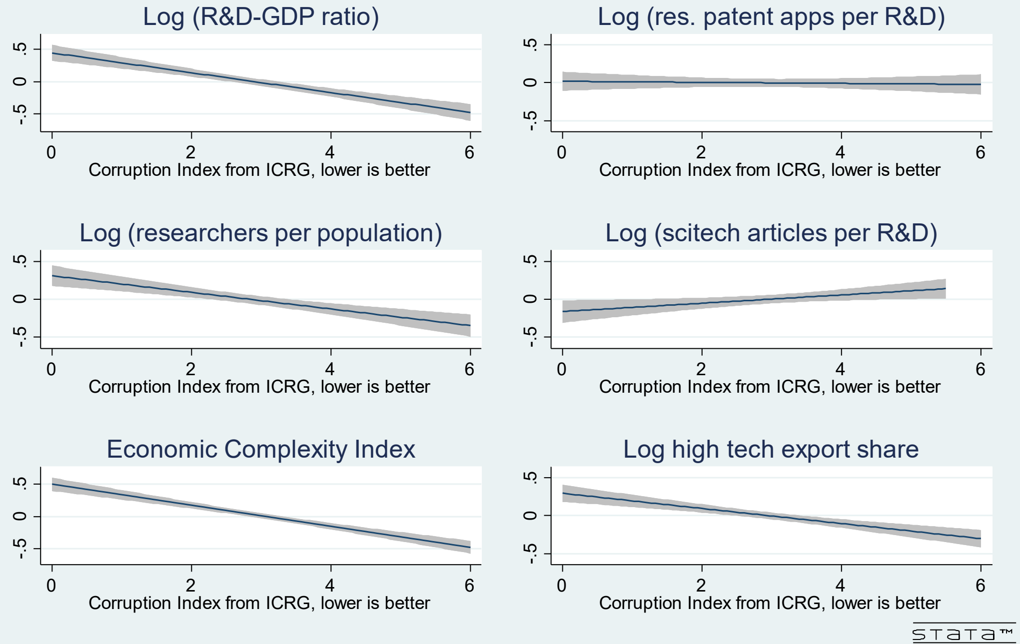
@Instruments are polity and GsoD variables (chosen out of 18).



Source: ICRG, PRS Group.

A few (6) observations have been rounded to the nearest integer or half integer in the all observation sample.

Figure 2: Correlation between the corruption index and standardized innovation variables



Appendices to How does expropriation risk affect innovation? (Not for publication)

Jose Miguel Benavente, Claudio Bravo-Ortega , Pablo Egaña-del Sol, and Bronwyn H. Hall

A. Data

Our data come from multiple sources, which we describe below. The variables themselves are described in Table A1. Table A4 provides a list of the 130 countries in our base sample along with indicators of the availability of the innovation data for these countries. Note that these data sources are of different vintages and the country names and codes they use vary slightly. For example, they are affected by the creation of South Sudan, the renaming of Eswatini from Swaziland, the breakup of Czechoslovakia, and so forth. We have standardized the names and codes across them and done our best to ensure that each country does not change its geographic dimension during our time period of 1996 to 2014, leading to unbalanced data in some cases.

1. The Polity IV Project (**Polity IV**), Center for Systemic Peace, Marshall et al. (2019)

This data source contains a number of variables on regime characteristics and political systems. It is a yearly unbalanced panel for 1800-2018 with 22 countries in 1800 increasing to 167 in 2018. We use data from 1996 to 2014 for which there are 146 countries with good data for at least some years.

2. International Country Risk Guide (**ICRG**), PRS Group, Howell (2020)

The source of the corruption index, which is based on a number of political risk factors. This is a balanced panel of 146 countries for 1984-2017.

3. Global State of Democracy (**GSoD**), International IDEA, Tufis (2018)

This source contains a number of political regime variables, variables describing barriers to participation and media freedom. It is an unbalanced panel of 157 countries and 30 broader regions for 1975-2017. We use only the country data and the years 1996-2014, for a total of 157 countries. A very few variable values are missing in some countries and some years.

4. **IMF** DATA

IMF data is our source of export diversity for the years 1996-2014. On their website, the export diversification measure is available for 200 countries from 1962 to 2014. http://data.imf.org

5. Atlas of Economic Complexity (**ECI**), Harvard Growth Laboratory.

From here we obtain the ECI, although our version of these data is somewhat older, containing more years (1962 to 2020) for 248 countries. https://atlas.cid.harvard.edu/

6. **UNESCO** (United Nations Educational, Scientific and Cultural Organization): World Development Indicators (**WDI**)

From here we obtain R&D and scientific publication variables from many countries, including developing (this data source is broader than OECD coverage, but much the same as the OECD for developed countries). We also get the high tech export share from these data.

7. **WIPO** (World Intellectual Property Organization)

This is our source for patent application statistics, by residents and non-residents over time. We use only the resident (domestic) applications.

Data sources 1-5 define our basic sample of 130 countries. For the base sample, we require good data on all the variables for at least 4 years per country. Most countries have the full 19 years, as can be seen in Table A4. Because the innovation outcome variables are more often missing, we use different samples for those regressions rather than lose too many observations.

Table A1

Sources of variables

Source	*Variable*	*description*	*Countries*	*Years*
UNESCO Inst. for statistics	rd_gdp	GERD as a percentage of GDP	149	1996-2018
	rd_ppp05	GERD in $1000 (PPP, in constant 2005 prices)	140	1996-2018
	rd_ppp_cur	GERD in $1000 (PPP, current prices)	141	1996-2018
	rd_gdp	GERD as a percentage of GDP	149	1996-2018
	rd_gov_ppp05	GERD financed by govt. (in $1000 PPP, in constant 2005 prices)	111	2012-2018
	rd_fte_ppp05	GERD per researcher, FTE (in $1000 PPP, constant 2005 prices)	122	1996-2018
	rd_fte_ppp_cur	GERD per researcher, FTE (in $1000 current PPP$)	122	1996-2018
	rd_hc_ppp05	GERD per researcher, HC (in $1000 PPP, constant 2005 prices)	134	1996-2018
	rd_hd_ppp_cur	GERD per researcher, HC (in $1000 current PPP$)	135	1996-2018
WDI	ex_hitech_share	Share of high-tech products in exports	129	1996-2018
WIPO IP Statistics Data Center	pat_noresid	Non-resident patent apps to country's office in year	171	1990-2020
	pat_resid	Resident patent apps to country's office in year	171	1990-2020
	pat_tot	Total patent apps to country's office in year	171	1990-2020
International Country Risk Guide ICRG)	corruption	index of demands for bribes, exchange controls, or tax assessments for those in intl. trade; patronage and nepotism; ties between politics and business	146	1996-2014
The Polity IV Project, Center for Systemic Peace	autoc	Institutionalized autocracy	167	1996-2018
	comp_exec	Competitiveness of chief executive recruitment	167	1996-2018
	democ	Institutionalized democracy	167	1996-2018
	open_exec	Openness of chief executive recruitment	167	1996-2018
	parcomp	Competitiveness of participation in govt	167	1996-2018
	parreg	Regulation of participation in govt.	167	1996-2018
	reg_exec	Regulation of chief executive recruitment	167	1996-2018
Global State of Democracy	access_Justice	Access to justice	157	1996-2014
	barriers_parties	Barriers to political parties	157	1996-2014
	clean_election	Freedom of elections from irregularities	157	1996-2014
	check_govern	Checks on government	157	1996-2014
	civil_Liberties	Civil liberties	157	1996-2014
	civil_part	Civil society participation	157	1996-2014
	comp_exec_g	Competitiveness of chief exec recruitment	157	1996-2014
	comp_part	Competitiveness of alternative policy participation	157	1996-2014
	court_indep	High court independence	157	1996-2014
	critical_print	Critical print and broadcast media	157	1996-2014
	harass_journ	Harassment of journalists	157	1996-2014
Atlas of Economic Complexity	eci	Economic complexity index	248	1962-2020
IMF	diversity	Index of export diversity	199	1996-2014

Abbreviations

GERD - Gross Domestic Expenditure on R&D

HC - headcount

Table A2

Correlation matrix for corruption and risk variables

Variables	*corruption*	*autoc*	*comp_exec_p*	*democ*	*open_exec*	*parcomp*	*parreg*	*reg_exec*	*access_justice*	*barriers_parties*	*clean_election*
corruption	1.000										
autoc	0.191	1.000									
comp_exec_p	0.219	0.772	1.000								
democ	0.250	0.865	0.914	1.000							
open_exec	0.107	0.502	0.735	0.568	1.000						
parcomp	0.213	0.825	0.718	0.849	0.472	1.000					
parreg	0.131	-0.154	0.046	0.164	-0.058	0.202	1.000				
reg_exec	0.227	0.429	0.773	0.687	0.337	0.510	0.187	1.000			
access_justice	0.281	0.534	0.561	0.695	0.315	0.686	0.456	0.506	1.000		
barriers_parties	0.187	0.810	0.710	0.823	0.439	0.782	0.105	0.480	0.662	1.000	
clean_election	0.241	0.768	0.766	0.852	0.450	0.758	0.181	0.607	0.773	0.776	1.000
check_govern	0.262	0.767	0.716	0.850	0.416	0.811	0.250	0.558	0.859	0.805	0.884
civil_liberties	0.248	0.823	0.734	0.861	0.453	0.844	0.179	0.538	0.833	0.861	0.875
civil_participation	0.262	0.761	0.658	0.762	0.384	0.752	0.143	0.496	0.764	0.777	0.805
comp_exec_g	0.231	0.833	0.969	0.942	0.585	0.747	0.061	0.761	0.579	0.762	0.797
comp_participation	0.246	0.751	0.711	0.859	0.430	0.821	0.308	0.575	0.703	0.775	0.772
court_indep	0.216	0.645	0.647	0.756	0.366	0.714	0.239	0.513	0.769	0.690	0.764
critical_media	0.221	0.798	0.679	0.803	0.405	0.794	0.097	0.491	0.694	0.807	0.814
harass_journ	0.277	0.713	0.663	0.802	0.404	0.768	0.296	0.495	0.839	0.768	0.828

	check_govern	*civil_liberties*	*civil_participation*	*comp_exec_g*	*comp_participation*	*court_indep*	*critical_media*	*harass_journ*
check_govern	1.000							
civil_liberties	0.925	1.000						
civil_participation	0.884	0.893	1.000					
comp_exec_g	0.754	0.779	0.704	1.000				
comp_participation	0.785	0.817	0.732	0.747	1.000			
court_indep	0.885	0.784	0.722	0.672	0.658	1.000		
critical_media	0.898	0.892	0.861	0.726	0.746	0.759	1.000	
harass_journ	0.896	0.921	0.854	0.702	0.784	0.742	0.830	1.000

Table A3

Correlation of Corruption variables and R&D/innovation variables

	corruption	*autoc*	*comp_exec_p*	*democ*	*open_exec*	*parcomp*	*parreg*	*reg_exec*	*access_justice*	*barriers_parties*	*clean_election*
Log R&D	-0.215	0.116	-0.060	-0.123	-0.077	-0.135	-0.312	-0.127	-0.382	-0.218	-0.108
Log GDP	-0.169	0.056	-0.102	-0.153	-0.094	-0.128	-0.248	-0.176	-0.289	-0.248	-0.123
Log R&D/GDP	-0.305	-0.020	-0.187	-0.297	-0.104	-0.367	-0.556	-0.210	-0.708	-0.315	-0.317
Log resid. patents	-0.168	0.151	-0.012	-0.049	-0.101	-0.038	-0.183	-0.061	-0.256	-0.124	-0.019
Log res. pats per R&D	-0.048	0.127	0.046	0.053	-0.085	0.084	0.034	0.037	-0.012	0.030	0.086
Log sci tech articles	-0.219	0.132	-0.055	-0.143	-0.078	-0.158	-0.372	-0.140	-0.399	-0.200	-0.119
Log aritcles per R&D	-0.043	0.080	0.012	-0.094	-0.012	-0.112	-0.279	-0.067	-0.113	0.048	-0.057
Log researchers/pop	-0.266	-0.037	-0.220	-0.356	-0.161	-0.419	-0.650	-0.230	-0.749	-0.326	-0.385
Log hitech export share	-0.167	0.065	-0.114	-0.123	-0.164	-0.079	-0.213	-0.067	-0.334	-0.066	-0.084
Diversity/complexity	-0.214	-0.011	-0.151	-0.279	-0.122	-0.258	-0.407	-0.211	-0.393	-0.323	-0.177

	check_govern	*civil_liberties*	*civil_participation*	*comp_exec_g*	*comp_participation*	*court_indep*	*critical_media*	*harass_journ*
Log R&D	-0.265	-0.197	-0.310	-0.050	-0.135	-0.207	-0.119	-0.357
Log GDP	-0.238	-0.192	-0.276	-0.092	-0.128	-0.207	-0.147	-0.303
Log R&D/GDP	-0.510	-0.433	-0.492	-0.182	-0.367	-0.380	-0.246	-0.616
Log resid. patents	-0.191	-0.096	-0.200	0.005	-0.038	-0.125	-0.059	-0.242
Log res. pats per R&D	-0.032	0.055	0.002	0.064	0.084	0.017	0.033	-0.016
Log sci tech articles	-0.291	-0.204	-0.279	-0.045	-0.158	-0.268	-0.127	-0.385
Log aritcles per R&D	-0.136	-0.049	0.091	0.015	-0.112	-0.270	-0.044	-0.156
Log researchers/pop	-0.561	-0.481	-0.450	-0.207	-0.419	-0.434	-0.266	-0.656
Log hitech export share	-0.210	-0.145	-0.238	-0.092	-0.079	-0.173	-0.005	-0.276
Diversity/complexity	-0.288	-0.291	-0.286	-0.139	-0.258	-0.342	-0.239	-0.443

Table A4

Countries used in estimation

		Number of years with		
Country	*Base sample*	*R&D data*	*R&D employee data*	*Export data*
Albania	19	2	0	19
Algeria	19	5	0	19
Angola*	19	0	0	0
Argentina	19	19	18	19
Armenia	19	18	0	17
Australia	19	18	15	19
Austria	19	19	14	19
Azerbaijan	19	19	0	19
Bangladesh	19	0	0	15
Belarus	19	19	0	17
Belgium	19	19	19	16
Bolivia	19	8	7	19
Botswana	19	3	2	15
Brazil	19	15	15	19
Bulgaria	19	19	19	19
Burkina Faso	19	12	0	18
Cameroon	19	0	0	17
Canada	19	19	19	19
Chile	19	8	8	19
China	19	19	19	19
Colombia	19	19	2	19
Congo	19	0	5	8
Congo, Dem. Rep.	9	4	0	0
Costa Rica	19	19	12	19
Croatia	19	16	17	19
Cuba	19	19	0	0
Cyprus	19	17	17	19
Czechia	19	19	19	19
Côte d'Ivoire	9	0	0	9
Denmark	19	19	19	19
Dominican Republic	19	0	0	16
Ecuador	19	19	19	19
Egypt	18	15	7	18
El Salvador	19	9	0	19
Eritrea*	19	0	0	0
Estonia	19	17	17	19
Finland	19	19	11	19
France	19	19	19	19
Gabon	19	3	0	14
Gambia	19	4	4	17
Germany	19	19	19	19
Ghana	19	4	4	17
Greece	19	18	15	19

Table A4 (cont.)

Guatemala	19	10	10	19
Guinea	19	0	0	14
Guinea-Bissau*	19	0	0	0
Haiti	11	0	0	2
Honduras	19	5	0	17
Hungary	19	19	19	19
India	19	19	7	19
Indonesia	19	4	3	19
Iran	19	13	9	12
Iraq	12	5	5	0
Ireland	19	19	19	19
Israel	19	19	0	19
Italy	19	19	19	19
Jamaica	19	2	0	15
Japan	19	19	19	19
Jordan	19	2	0	18
Kazakhstan	19	18	8	19
Kenya	19	4	4	16
Korea, Dem. Rep.*	19	0	0	0
Korea, Republic of	19	19	19	19
Kuwait	19	18	18	0
Latvia	19	19	19	19
Lebanon*	10	0	0	0
Liberia*	15	0	0	0
Libya*	15	0	0	0
Lithuania	19	19	19	19
Madagascar	19	18	14	19
Malawi	19	0	4	19
Malaysia	19	19	19	19
Mali	19	4	2	16
Mexico	19	19	19	19
Moldova	19	14	12	19
Mongolia	19	18	0	0
Morocco	19	10	9	19
Mozambique	19	6	5	18
Myanmar	19	6	6	0
Namibia	19	2	0	0
Netherlands	19	19	19	19
New Zealand	19	17	18	19
Nicaragua	19	6	0	19
Niger	19	0	6	19
Nigeria	19	0	0	17
Norway	19	18	18	19
Oman	19	4	4	0
Pakistan	19	17	11	19
Panama	19	19	18	19
Papua New Guinea*	19	0	0	0
Paraguay	19	14	14	19

Table A4 (cont.)				
Peru	19	12	0	19
Philippines	19	12	13	19
Poland	19	19	19	19
Portugal	19	19	19	19
Qatar	19	0	3	0
Romania	19	19	19	19
Russian Federation	19	19	19	19
Saudi Arabia	19	11	0	0
Senegal	19	3	5	19
Serbia	9	9	8	9
Sierra Leone	14	0	0	2
Singapore	19	19	19	19
Slovakia	19	19	19	19
Slovenia	19	19	19	19
South Africa	19	15	15	19
Spain	19	19	19	19
Sri Lanka	19	13	13	16
Sudan	19	7	0	0
Sweden	19	18	18	19
Switzerland	19	5	7	19
Syrian Arab Republic	19	0	0	8
Taiwan*	19	0	0	0
Tanzania	19	7	4	18
Thailand	19	19	19	19
Togo	19	5	12	18
Trinidad and Tobago	19	19	0	15
Tunisia	16	10	5	16
Türkiye	19	19	19	19
Uganda	19	10	2	19
Ukraine	19	18	9	19
United Arab Emirates	19	4	0	0
United Kingdom	19	19	19	19
United States	19	19	19	19
Uruguay	19	16	11	19
Venezuela	19	10	17	17
Viet Nam	19	4	3	0
Yemen*	19	0	0	0
Zambia	19	9	5	19
Zimbabwe	19	0	0	17
Total	**2399**	**1446**	**1134**	**1882**

* Country has no innovation data other than ECI.

B. Additional regression tables

Table B1

Country OLS and FE estimates

Dependent variable:	*Log R&D spending*	*Log R&D/ GDP*	*Log resident pat apps*	*Log resident pat apps per R&D*	*Log Sci & Tech articles*	*Log aritcles per R&D*	*Log R&D Research-ers*	*Log R&D Research-ers per pop.*	*Log Hi tech export share*	*Economic complexity index*
					Pooled OLS					
Corruption	-0.417***	-0.180***	-0.421***	-0.004	-0.258**	0.044	-0.190	-0.196**	-0.223***	-0.170***
	(0.121)	(0.055)	(0.144)	(0.063)	(0.124)	(0.048)	(0.119)	(0.091)	(0.078)	(0.041)
Standard error	2.598	1.140	3.208	1.863	2.308	0.855	2.122	1.571	2.122	1.017
R-squared	0.047	0.043	0.033	0.007	0.032	0.027	0.019	0.033	0.022	0.045
					OLS with Country effects					
Corruption	-0.028**	-0.001	-0.024	0.004	-0.021**	0.013	0.001	0.002	-0.014	0.009
	(0.013)	(0.013)	(0.037)	(0.036)	(0.011)	(0.016)	(0.013)	(0.012)	(0.031)	(0.012)
Standard error	0.314	0.303	0.879	0.893	0.214	0.324	0.260	0.246	0.789	0.258
R-squared	0.004	0.000	0.001	0.006	0.013	0.023	0.001	0.002	0.006	0.011
Observations	1446	1446	1446	1446	1199	1199	1134	1134	1882	2391
N of countries	106	106	106	106	106	106	88	88	107	130

Year fixed effects included in all regressions.
Standard errors are robust, clustered on country.
*** p<0.01, ** p<0.05, * p<0.1

Table B2
First stage estimates

	Without year effects			*With year effects*			*Selected variables only*		
Variable	*Coeff.*	*s.e.*	*p-value*	*Coeff.*	*s.e.*	*p-value*	*Coeff.*	*s.e.*	*p-value*
Autocracy	0.108	0.171	0.528	0.138	0.217	0.528			
Chief exec comp. (Polity)	-0.576	0.447	0.201	-1.816	1.416	0.202			
Democracy	0.147	0.172	0.394	0.080	0.259	0.757			
Chief exec openness	0.145	0.135	0.284	0.494	0.430	0.253			
Gov participation competitive	-0.025	0.085	0.772	0.013	0.137	0.926			
Gov participation regulated	0.006	0.059	0.913	*-0.125*	*0.123*	*0.312*			
Chief exec choice regulated	***0.184***	***0.098***	***0.063***	0.099	0.106	0.352	***0.131***	***0.069***	***0.023***
Access to justice	***1.106***	***0.523***	***0.037***	1.706	1.351	0.209	0.963	1.134	0.397
Barriers to parties	-0.366	0.301	0.227	-0.337	0.445	0.450			
Clean elections	-0.175	0.353	0.621	-0.159	0.372	0.671			
Checks on Gov	0.010	0.864	0.991	*-2.132*	*1.149*	*0.066*			
Civil liberties	***-1.441***	***0.747***	***0.056***	0.302	1.621	0.853			
Civil society participation	0.723	0.494	0.146	***1.868***	***0.839***	***0.028***	***1.424***	***0.621***	***0.023***
Chief exec comp. (GsoD)	0.819	0.852	0.338	3.476	2.775	0.213			
Alt policy competitiveness	0.015	0.291	0.960	-0.378	0.518	0.467			
High court independence	-0.129	0.263	0.624	-0.063	0.432	0.885			
Critical media	0.006	0.480	0.991	0.078	0.723	0.914			
Harassment of journalists	***0.972***	***0.580***	***0.096***	0.271	0.790	0.732	-0.032	0.683	0.963
Country effects		no			yes			yes	
Year effects		yes			yes			yes	
R-squared		0.119			0.053			0.098	
Standard error		0.888			0.776			0.780	

2,399 observations

Table B3

Country IV and IVFE estimates

	Log R&D spending	*Log R&D/ GDP*	*Log resident pat apps*	*Log resident pat apps per R&D*	*Log Sci & Tech articles*	*Log aritcles per R&D*	*Log R&D Research-ers*	*Log R&D Research-ers per pop.*	*Log Hi tech export share*	*Economic complexi ty index*
					Pooled IV regression					
Corruption	-4.362***	-2.119***	-4.091***	0.149	-4.091***	0.728***	-1.881***	-2.405***	-2.818***	-1.993***
	(0.824)	(0.340)	(0.890)	(0.189)	(0.901)	(0.226)	(0.624)	(0.551)	(0.613)	(0.300)
Standard error	4.436	2.113	4.620	1.858	4.046	1.039	2.618	2.582	3.178	1.933
LM test for weak inst.	19.80	19.80	19.80	19.80	15.08	15.08	17.26	17.26	27.16	30.93
D. F. for weak inst test	18	18	18	18	18	18	18	18	18	18
p-value for LM test	0.344	0.344	0.344	0.344	0.656	0.656	0.505	0.505	*0.076*	***0.029***
Hansen (J) for overid.	13.45	14.98	21.30	20.92	11.97	24.14	14.69	24.39	12.11	16.41
D. F. for J test	17	17	17	17	17	17	17	17	17	17
p-value for J test	0.706	0.597	0.213	0.230	0.802	0.116	0.618	0.109	0.794	0.495
					IV with country effects					
Corruption	-0.346	-0.225	-0.111	0.235	-0.271	0.261	-0.062	-0.059	0.020	0.052
	(0.235)	(0.228)	(0.267)	(0.400)	(0.224)	(0.302)	(0.170)	(0.149)	(0.257)	(0.108)
Standard error	0.401	0.357	0.908	0.938	0.277	0.376	0.273	0.258	0.808	0.266
LM test for weak inst.	28.14	28.14	28.14	28.14	19.00	19.00	17.36	17.36	19.72	22.98
D. F. for weak inst test	18	18	18	18	18	18	17	17	18	18
p-value for LM test	*0.060*	*0.060*	*0.060*	*0.060*	0.392	0.392	0.430	0.430	0.349	0.192
Hansen (J) for overid.	12.13	18.38	14.13	8.66	19.26	13.78	19.32	18.48	12.17	14.76
D. F. for J test	17	17	17	17	17	17	16	16	17	17
p-value for J test	0.792	0.365	0.658	0.950	0.314	0.683	0.253	0.297	0.790	0.612
Pbservations	1446	1446	1446	1446	1199	1199	1134	1134	1882	2391
N of countries	106	106	106	106	106	106	88	88	107	130

Year fixed effects included in all regressions.

Standard errors are robust, clustered on country.

*** p<0.01, ** p<0.05, * p<0.1

Excluded instruments are the full set of Polity and GSoD variables.