Interaction Effects in Econometrics

October 16, 2011

Abstract

We provide practical advice for applied economists regarding robust specification and interpretation of linear regression models with interaction terms. We replicate a number of prominent published results using interaction effects and examine if they are robust to reasonable specification permutations.

JEL classification: C12, C13
Keywords: Non-Linear Regression, Interaction Terms.

1 Introduction

A country may consider a reform that would strengthen the financial sector. Would this help economic growth and development? This simple question is frustratingly hard to answer using empirical data because economic development itself spawns financial development, so while economic and financial developments are positively correlated this does not answer the question asked. In a highly influential paper, Rajan and Zingales (1998) provide convincing evidence that financial development is important for economic development by asking the simple question: do industrial sectors that are more dependent on external finance grow faster in countries with a high level of development. This question involves *interactions* between financial development and dependency on external finance. Since the publication of Rajan and Zingales' highly influential study, the estimation of models with interaction effects have become very common in applied economics.

Many articles applying interaction terms are motivated in an intuitive fashion similar to the story we just outlined which makes robustness analysis particularly important. Robustness analysis with respect to variables included besides the main variable(s) of interest are now routinely performed in most empirical articles. However, it is our view that robustness analysis with respect to the functional form should be standard when one uses non-linear specifications, in particular those involving interactions which we focus on here.[1] This article discusses the case where the true specification, which we will also refer to as the true Data Generating Process (DGP), is not precisely known.

If the DGP is not pinned down by theory, the standard linear specification can be seen as a first order Taylor series expansion. If one wants to examine the role of, say, $x * z$ in a relation $y = f(x, z)$, then we argue that it is reasonable to examine if the interaction term may be picking up other left-out components in the second order expansion of f. Further, we believe that it is often informative to consider the interaction of, say,

[1] In the case where the specification of the empirical model is tightly pinned down by theory, "robustness analysis" is rather a test of the underlying theory.

x and z after these have been transformed to be orthogonal to other variables. For example, if z is, say, financial openness, and one is interested in how financial openness affects the impact of x on y one would include $x * z$. But z may be correlated with other variables and including linear terms of those will not prevent $x * z$ from spuriously picking up the affect of the interaction of some of those variables with x. However, if one orthogonalizes z to other variables, this will not happen. This is, of course, what OLS does automatically in a linear regression but in non-linear specifications the researcher needs to explicitly consider this. In our experience, this form of robustness analysis is rarely performed and this article calls for this to become as standard as other typical robustness checks. We further point out a few issues of interpretation and very briefly discuss choice of instruments when interaction terms are included.

We replicate parts of five influential articles, starting with Rajan and Zingales (1998), checking if their results are robust. The second paper is also written by Rajan and Zingales (2003) who examined if the number of listed firms in a country is affected by openness and the historical (1913) level of industrialization. The third article, by Castro, Clementi, and MacDonald (2004), hypothesizes that strengthening of property rights is beneficial for growth and more so when restrictions on capital transactions (capital flows) are weaker, while the fourth article, Caprio, Laeven, and Levine (2007), examines if bank valuations (relative to book values) are higher where owners have stronger rights. The fifth and last replication examines Spilimbergo (2009) who studies if countries that send large number of students abroad have better democracies. We find that most of these papers, if not Spilimbergo's, are robust to our suggested robustness tests and for several of these, some of our alternative specifications strengthen the authors' cases. The specification tested in the Spilimbergo article is one of many specifications that he employs and our results for this case are better seen as illustrating our suggestions than as a serious criticism of the conclusions of the article.

In Section 2, we discuss some practical issues related to the specification of regressions with interaction effects, illustrate our recommendations with Monte Carlo simulations, and make recommendations for practitioners. In Section 3, we revisit some prominent applied papers where interaction effects figure prominently, including Rajan and Zingales (1998), and examine if the published results are robust and Section 4 concludes.

2 Linear Regression with Interaction Effects

Many econometric issues related to models with interaction effects are very simple and we illustrate our discussion using simple Ordinary Least Squares (OLS) estimation. Often applied papers use more complicated methods involving, say, Generalized Method of Moments, clustered standards errors, etc., but the points we are making typically carry over to such settings with little modification.

Let Y be a dependent variable, such as growth of an industrial sector, and X_1 and X_2 independent variables that may impact on growth, such as the dependency on external finance and financial development. Applied econometricians have typically allowed for interaction effects between two independent variables, X_1 and X_2 by estimating a simple multiple regression model of the form:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 \, X_2 + \epsilon \,, \tag{1}$$

where $X_1 \, X_2$ refers to a variable calculated as the simple observation-by-observation product of X_1 and X_2. In the example of Rajan and Zingales (1998), the interest centers around the coefficient β_3—a significant positive coefficient implies that sectors that are more dependent on external finance grow faster following financial development. We refer to the independent terms X_1 and X_2 as *"main terms"* and the product of the main terms, $X_1 X_2$, as the *"interaction term."* This brings us to our first simple observations.

2.1 Interpreting the t-statistics on the main terms

1. The partial derivative of Y with respect to X_1 is $\beta_1 + \beta_3 X_2$. The interpretation of β_1 is the partial derivative of Y with respect to X_1 when $X_2 = 0$. A t-test for $\beta_1 = 0$ is, therefore a test of the null of no effect of X_1 when $X_2 = 0$. To test for no effect of X_1 one needs to test if $(\beta_1, \beta_3) = (0, 0)$ using, for example, an F-test.

In applied papers, the non-interacted regression

$$Y = \lambda_0 + \lambda_1 X_1 + \lambda_2 X_2 + \upsilon, \tag{2}$$

is often estimated before the interacted regression. In this regression, $\lambda_1 = \partial Y / \partial X_1$ is the partial derivative of Y with respect to X_1, implicitly evaluated at $X_2 = \overline{X}_2$ (the mean value of X_2).[2] The estimated β_1-coefficient in (1) is typically very close to $\hat{\lambda}_1 - \hat{\beta}_3 \overline{X}_2$.

2. Estimating the interacted regression in the form

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 - \overline{X}_1)\,(X_2 - \overline{X}_2) + \epsilon\,, \tag{3}$$

results in the exact same fit as equation (1) and the exact same coefficient $\hat{\beta}_3$ and is nothing but a renormalization. $\hat{\beta}_1$ will typically be close to $\hat{\lambda}_1$ estimated from equation (2) because $\beta_1 = \partial Y / \partial X_1$ is the partial derivative of Y with respect to X_1, evaluated at $X_2 = \overline{X}_2$. If a researcher reports results from (2) and wants to keep the interpretation of the coefficient to the main terms similar, it is usually preferable to report results of the regression (3) with demeaned interaction terms even if it is the exact same statistical model in a different parameterization.[3]

[2] Some social scientists suggest that the interaction term undermines the interpretation of the regression coefficients associated with X_1 and X_2 (e.g., Allison (1977), Althauser (1971), Smith and Sasaki (1979), and Braumoeller (2004)). The point is simply that researchers sometimes do not notice the change in the interpretation of the coefficient estimate for the main terms when the interaction term is added.

[3] Because $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 - \overline{X}_1)(X_2 - \overline{X}_2) = (\beta_0 + \beta_3 \overline{X}_1 \overline{X}_2) + (\beta_1 - \beta_3 \overline{X}_2)\, X_1 + (\beta_2 - \beta_3 \overline{X}_1) X_2 + \beta_3 X_1\, X_2$, we get the exact same fit, with the changes in the estimated parameters given

2.1.1 Monte Carlo simulation

We first illustrate how the specification of the interaction term affects the interpretation of the main terms, although we are not the first to make this point. We generate a dependent variable, Y, as $Y = 3\ X_1 + 5\ X_2 + 8\ X_1 X_2 + \epsilon$, where $X_1 = 1 + \epsilon_1$ and $X_2 = 1 + \epsilon_2$, $\epsilon_i \sim N(0,1)$, for all i. We estimate model (2) without an interaction term (that model is misspecified) because it is often natural to start by estimating equation (2) when it is not priori obvious if an interaction effect should be included. Next, we allow for an interaction term that is either demeaned or not. The latter specifications are both correctly specified. In column (1) of Table 1, the results for the model without an interaction term are presented and, in columns (2) and (3), the correctly specified model is estimated. In column (2), we see how the coefficient to X_1 changes from about 11 to about 3 when the regressors are not demeaned before they are interacted—a change is close to the predicted size of $\beta_3 E\{X_2\}$. The large change in the coefficient to the main term is not due to misspecification but it reflects that the coefficient to X_1 is to be interpreted as the marginal effect of X_1 when X_2 is zero. In column (3), we estimate model (3) where the terms in the interaction are demeaned and the coefficient to interaction term is unchanged from column (2) while the coefficients of main terms are very close to the ones in column (1)—with the same interpretation.

2.2 A simple observation on IV estimation

3. In the case where, say, X_2 is endogenous, X_1 is exogenous, and Z is a valid instrument for X_2, $X_1 Z$ will be a valid instrument for $X_1 X_2$.

from the correspondence between the left- and right-hand side of this equality. E.g., $\hat{\lambda}_0$ will be equal to $\hat{\beta}_0 + \beta_3 \overline{X}_1 \overline{X}_2$.

2.3 Robustness to misspecification

If one considers second order terms, a more general specification that one may want to consider for robustness, is the full second order expansion

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 - \overline{X}_1)\,(X_2 - \overline{X}_2) + \beta_4 X_1^2 + \beta_5 X_2^2 \; + \epsilon\,. \tag{4}$$

(We will refer to $X_i^2; i = 1, 2$ as "second-order terms"—in applications one may wish to enter the second-order terms in a demeaned forms for the same reasons as discussed for the interaction term but for notational brevity we use the simpler non-demeaned form here.) The relevance of this observation is as follows.

4. In a regression with interaction terms, the main terms should always be included unless excluded by economic theory. Otherwise, the interaction effect may be significant due to left-out variable bias. ($X_1\,X_2$ is by construction likely to be correlated with the main terms.)[4]

5. If $Y = f(X_1, X_2)$ can be approximated by the second order expansion (4) with a non-zero coefficient to either X_1^2 or X_2^2 and $corr(X_1, X_2) \neq 0$, the coefficient β_3 in the interacted regression (1) may be spuriously significant. For example, if $corr(X_1, X_2) > 0$ the estimated coefficient $\hat{\beta}_3$ will usually be positive even if $\beta_3 = 0$. If quadratic terms are not otherwise ruled out, we recommend also estimating the specification (4) in order to verify that a purported interaction term is not spuriously capturing left-out squared terms.

The potential bias from leaving out second order terms is easily understood. If X_1 and X_2 are (positively) correlated, we can write $X_2 = \alpha X_1 + w$ (where α is positive) so

[4] Some authors have referred to this as a multicollinearity problem. Althauser (1971) shows that the main terms and the interaction term in the equation (1) are correlated. These correlations are affected in part by the size and the difference in the sample means of X_1 and X_2. Smith and Sasaki (1979) also argue that the inclusion of the interaction term might cause a multicollinearity problem. In our view, collinearity is not a problem for regressions with interaction effects of a different nature than elsewhere in empirical economics—if one asks too much from a small sample, correlations between regressors make for fragile inference.

the interaction term (we suppress the mean for simplicity) becomes $\alpha X_1^2 + X_1 w$ where the latter term has mean zero and will be part of the error in the regression. If X_1^2 is part of the correctly specified regression with coefficient δ, the estimated coefficient to the interaction term when estimating equation (1) will be $\alpha\,\delta$.

2.3.1 Monte Carlo simulation

In Table 2, the true model does not include an interaction term, instead it is nonlinear in one of the main terms. We simulate $Y = X_1 + X_1^2 + \epsilon$ where $X_1 = 1 + \epsilon_1$ and $X_2 = 1 + X_1 + \epsilon_2$, $\epsilon_i \sim N(0,1)$ for all i. When $corr(X_1, X_2) \neq 0$, as in this example, the interaction term might pick up a left-out variable effect. In column (1), we show the correct specification. In column (2), we estimate the interaction model and observe that the interaction term is highly significant. Our suggestion, to hedge against such spurious inference, is to include the squares of both main terms together with the interaction term. We report this specification in column (3). This model is correctly specified, albeit some regressors have true coefficients of zero and we get the correct result.

2.4 Panel data

Consider a panel data regression with left-hand side variable Y_{it} where i typically is a cross-sectional index, such as an individual or a country (we will use the term country, for brevity), and t a time index. Denote, for a generic panel data variable X_{it}, the average over time for cross-sectional unit i by $\overline{X}_{i.}$ (i.e., $\frac{1}{T}\Sigma_{t=1}^{T} X_{it}$), the average across cross-sectional units at period t by $\overline{X}_{.t}$, and the mean across all observations by $\overline{X}_{..}$.

Consider the panel data regression

$$Y_{it} = \mu_i + \nu_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 (X_{1it} - \overline{X}_{1..})\,(X_{2it} - \overline{X}_{2..}) + \epsilon_{it}\,, \tag{5}$$

where μ_i and ν_t are country- and time-fixed effects.

The regression (5) is not robust to squared terms as in the simple OLS case, but in the panel data situation this regression is also not robust to slopes that vary across, say, countries. If the correct specification is, say,

$$Y_{it} = \mu_i + \nu_t + \beta_1 X_{1it} + \beta_{i2} X_{2it} + \epsilon_{it} \,, \tag{6}$$

then, if the mean of X_1 varies by country and the covariance of $\overline{X}_{1i.}$ and β_{i2} is non-zero, the covariance of $(X_{1it} - \overline{X}_{1..})\,(X_{2it} - \overline{X}_{2..})$ and $\beta_{i2} X_{2it}$ becomes non-zero and the interaction term will pick up the country-varying slopes.

6. In order to hedge against the interacted regression (5) spuriously capturing country-varying slopes, we suggest that panel data regressions are estimated as

$$Y_{it} = \mu_i + \nu_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \beta_3 (X_{1it} - \overline{X}_{1i.})\,(X_{2it} - \overline{X}_{2i.}) + \epsilon_{it} \,,$$

   where the country-specific means are subtracted from each variable in the interaction. Of course, if the time-series dimension of the data is large, one may directly allow for country-varying slopes. This specification is suggested, in particular, in datasets where one may expect heterogeneity across the cross-sectional observations. Alternatively, this specification provides a useful robustness test.

Note that the panel data regression $Y_{it} = \mu_i + \nu_t + \beta_1 X_{1it} + \beta_2 X_{2it} + \epsilon_{it}$ is equivalent to the regression

$$Y_{it} = \beta_1 (X_{1it} - \overline{X}_{1..}) + \beta_2 (X_{2it} - \overline{X}_{2..}) + \epsilon_{it} \,, \tag{7}$$

and, indeed, that is how most software packages perform the estimation since this avoids having a large dimensional regressor matrix in case the cross-sectional or time dimension is large. This follows from the fact that a regression on a, say, country dummy is equivalent to subtracting the country-specific average and a simple application of the Frisch-Waugh theorem.[5]

[5] Frisch-Waugh (1933) theorem: Consider an equation $Y = X_1 \beta_1 + X_2 \beta_2 + \epsilon$ where β_1 is $k_1 \times 1$, β_2

2.4.1 Monte Carlo simulations: Panel data with varying slopes

We consider a panel data regression with two "countries" $i = 1, 2$ for $T = 500$ "years." The true model has the slope for X_2 varying across countries: $Y_{it} = \alpha_i + X_{1it} + \xi_i\, X_{2it} + \epsilon_{it}$.[6]

In Table 3, column (1) shows the results of estimating model (5). We find a spuriously significant coefficient to the interaction term and a coefficient to X_2 which is similar to the average of the true country-varying slopes. The variable X_1 has a lower mean for country 2 and since the slope of X_2 is higher for country 2 the least squares algorithm can minimize the squared errors by assigning a negative coefficient to the interaction term. In effect, the estimated model allows for different slopes to X_2 since $\partial Y_{it}/\partial X_2 = \beta_2 + \beta_3(X_{1it} - \overline{X}_{1..})$ which averaged over t for given i is different from 0. This is not the true model, but since the model estimated does not allow the slope to vary in any other way, this outcome occurs. In the second column, we illustrate how the simple suggestion of subtracting the country-specific means from each variable prevents the interaction term from becoming spuriously significant due to country-varying slopes.

2.5 Orthogonalizing the Regressors

In a situation where the regression of interest utilizes a large number of regressors, the estimated interaction term may capture all sorts of interactions between the variables. In this situation, one might ascertain that a regression with interactions really captures only interactions between innovations to the variables of interest by orthogonalizing the variables using the Frisch-Waugh theorem. Consider equation (1). If we want to find

is $k_2 \times 1$. The estimated coefficients to X_1 from an OLS regression of Y on X_1 and X_2 are identical to the set of coefficients obtained when the residuals from regressing Y on X_2 is regressed on the set of residuals from regressing X_1 on X_2. I.e., the OLS estimate of β_1 is $\hat{\beta}_1 = (X_1^{\psi\prime} X_1^{\psi})^{-1} X_1^{\psi\prime} Y^{\psi}$ where ($X_1^{\psi} = M_2\, X_1,\; Y^{\psi} = M_2\, Y,\; M_2 = [I - P_{X_2}]$ (M_2 is the residual maker from regressing X_1 on X_2), and $P_{X_2} = X_2\, ({X_2}' X_2)^{-1} {X_2}'$. This method is called "*netting out*" (or *partialing out*) the effect of X_2. Because we remove the linear effects of X_2, the cleaned variables Y^{ψ} and X_1^{ψ} are uncorrelated with ("orthogonal to") X_2.

[6] We set $X_{11t} = 1 + \epsilon_{1t}$ and $X_{21t} = 1 + X_{11t} + \epsilon_{2t}$ for the first country, $X_{12t} = 1/4 + \epsilon_{3t}$ and $X_{22t} = 1 + X_{12t} + \epsilon_{4t}$ for the second country where $\epsilon_{it} \sim N(0, 1)$ for all i. We allow the slope of X_2 to vary by country by setting $\xi_1 = 1$ and $\xi_2 = 2$.

the effect of X_1 on $\partial Y / \partial X_2$ and we want to ascertain that we are not picking up any other interaction or square term, we can interact X_2 with the Frisch-Waugh residual.

Case 1: If the concern is how the variable X_1, cleaned of any other regressors, affects the impact of X_2 on Y—or robustness with respect to this—we suggest running the following regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^{\psi}(X_2 - \overline{X_2}) + \beta_4 Z + \epsilon\,, \tag{8}$$

where Z is a third regressor (or vector of regressors), $X_1^{\psi} = M_2\, X_1$ and M_2 is the residual maker (from regressing X_1 on a constant, X_2, and Z). X_1^{ψ} is X_1 orthogonalized with respect to the other regressors.

If $\alpha_0 + \alpha_1 X_2 + \alpha_2 Z$ is the projection of X_1 on the other regressors, then $X_1\, X_2 = X_1^{\psi} X_2 + (\alpha_0 + \alpha_1 X_2 + \alpha_2 Z) * X_2$, which clearly illustrates how an effect of, say, $Z X_2$ on Y could make $X_1 X_2$ significant in the case where $\alpha_2 \neq 0$. Alternatively, the researcher could include X_2^2 and $Z * X_2$ in the regression; however, orthogonalization may be more convenient if the number of regressors is large relative to the sample size. Of course, it may be that the DGP is such that $X_1 X_2$ belongs in the regression, rather than $X_1^{\psi} X_2$. In either event, this robustness test can alert the econometrician to potential mis-specification.

Notice that this generalizes the subtraction of the average (equivalent to a regression on a constant) and the subtraction of "country-specific" averages. This procedure may not result in an unbiased coefficient to the interaction if it is truly the interaction of the non-orthogonalized X_1 and X_2 that affects Y; however, if the interaction involving orthogonalized terms is significant, it makes it less likely that the interaction is spurious. In either event, this robustness exercise may help the researcher obtain a better understanding of the data.

Case 2: If one wants to ascertain that the interaction of X_1 and X_2 captures no other regressors, a simple robustness check is to run the following regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^{\psi} X_2^{\psi} + \epsilon\,, \tag{9}$$

where $X_1^{\psi} = M_2\, X_1$ and $X_2^{\psi} = M_1\, X_2$, $M_1 = [I - P_{\beta_0, X_1}]$ and $M_2 = [I - P_{\beta_0, X_2}]$ (M_1 is the residual maker from regressing X_2 on a constant and X_1 and M_2 is the residual maker from regressing X_1 on a constant and X_2). In the case of other regressors in the specification, we suggest taking the residuals from a regression of all regressors. This specification does not deliver a consistent estimate for the coefficient to the interaction term if the DGP actually involves $X_1 * X_2$, but may alert the econometrician to potential problems if the specification is not tightly pinned down by theory.

2.6 Monte Carlo simulations: Frisch-Waugh orthogonalization

In Table 4, we simulate a model with an interaction term and correlated regressors and estimate various specifications and robustness regressions as suggested above. The first columns show the linear regression and a regression involving the demeaned interaction and quadratic terms. Of more interest is column (3) which uses the interaction of Frisch-Waugh orthogonalized terms. Orthogonalizing either X_1 or X_2, but not both, results in consistent estimates in this case. In column (6), using the interaction of orthogonalized terms, results in the quadratic terms in X_1 and X_2 being significant. Orthogonalizing either X_1 or X_2 leads to a consistent estimate for the interaction and non-zero quadratic terms for X_2 and X_1, respectively. A researcher doing a specification search will conclude that an interaction term belongs in the model but would need theoretical consideration to decide if quadratic terms should be included in a "best" specification.

Table 5 is an example where there is a significant interaction between X_1 and X_2 but the data generating process involves an interaction between X_1 and the part of X_2 that is orthogonal to X_1. In non-structural applications, it is often not obvious that whether the derivative of Y with respect to X_2 is a function of some X_1 or some variable which is part of X_1. For example, if the effect of credit varies by industry, it may not be "industry" (the type of product made) that matters but the correlation of industry dummies with financial structure. In the example here, the regressions where X_2 is Frisch-Waugh orthogonalized deliver consistent estimates while the regular interaction, still significant, does not—neither does the specification with both terms orthogonalized. The specification in column (5) is the true model. An investigator searching for specifications would notice the high t-value for the interaction term in this specification. The quadratic term in X_2 in column (7) is also very significant so an investigator would need to invoke theoretical considerations to choose between specifications—our suggested robustness tests do not substitute for this. They do, however, flag potential issues which the practice of reporting only a regression with X_1, X_2, and $X_1 * X_2$ does not.

Table 6 simulates a model with a data generating process which is quadratic in X_1 while X_1 and X_2 are correlated. In this case, the interaction term will be spuriously significant unless quadratic terms are included or either X_2 or both of the independent variables have been orthogonalized. Our suggestion is to include quadratic terms but this may be impractical in the case of a large amount of regressors in which case the orthogonalized regressions may be substitutes.

3 Replications

We replicate five important papers and examine if their implementation of interaction effects are robust. (Data details are given in the appendix.) First, in Table 7, we examine if the results of the highly influential paper of Rajan and Zingales (1998) are robust. The conclusion of this paper, which has by early 2010 has almost 2500 references, is that accounting standards matters—in particular in industries that are highly dependent on finance. Considering the influence of the paper, it is important to examine if the results are robust. The interactions of interest are between sectors' external financial dependence (E) and the country-level indicators of finance availability: the ratio of (total equity market capitalization plus domestic credit) to GDP (T) and accounting standards (A). We examine if the results are robust to using Frisch-Waugh residuals for total capitalization (T) and accounting standards (A). We find that using Frisch-Waugh residuals in the interaction term strengthens the size and significance of the interactions; in fact, the interaction of external dependence and equity market capitalization and credit turns from insignificant to clearly significant at the 5-percent level with the expected sign. Our robustness exercise makes the original claims of Rajan and Zingales (1998) empirically more convincing.

We also briefly consider the results of Rajan and Zingales (2003), who examined if the number of listed firms in a country is affected by openness (O), the historical (1913) level of industrialization (I), and the interaction of openness and historical industrialization. From Table 8, we see that the t-statistic on the main terms are very much affected by the interaction terms not being centered although this only involves a different interpretation of the t-statistic which are positive and significant when the variables are centered in the interaction. The impact of the interaction term is very robust to including quadratic terms in the main variables or orthogonalizing the regressors (indicating that these were likely to be near-orthogonal to begin with). Overall, the conclusions of Rajan and Zingales (2003) are robust to the potential misspecifications that we suggest examining.

Castro, Clementi, and MacDonald (2004) hypothesize that strengthening of property rights, as measured by laws mandating "one share-one vote," "anti-director rights" (which limit the power of directors to extract surplus), "creditor rights," and "rule of law," are beneficial for growth and more so when restrictions on capital transactions (capital flows) are weaker. They examine this by including interaction terms between the property rights indices and capital restrictions. Table 9 replicates Table 1 of their paper. We do not display regressions with centered interactions because the interaction terms are the variables of interest. In column (2), quadratic terms for the property rights measures are included but this strengthens the authors' main result of negative interactions. In column (3), we include a quadratic term in log GDP, to which weakens the significance of the parameters of interest below standard significance but we do not further explore this issue which is not at the focus of this article.[7] If we use Frisch-Waugh residuals for either the creditor rights measures of the capital restrictions measure we again find that the estimated interactions are mainly negative. Overall, the point estimates in the Castro, Clementi, and MacDonald (2004) study are not all robust, as one might conjecture from the size of the t-statistics, but the overall message of their regressions appears very robust to the kind of robustness checks we recommend.

Caprio, Laeven, and Levine (2007) examine if bank valuations (relative to book values) are higher where owners have stronger rights ($Rights$), as measured by an anti-director index, and whether this result is stronger when a larger share of cash flows (CF) accrues to the owners. The first column of Table 10 replicates Table 5 (column 1) of Caprio, Laeven, and Levine (2007). Column (2) includes quadratic terms and centers the variables before interacting. The very large t-statistic found for the main term, "rights", in column 1 turns insignificant and both main variables change signs. The non-centered implementation of Caprio, Laeven, and Levine (2007), in our opinion, gives a potentially misleading impression of the effect of the main terms; for example, the t-statistic of "rights" in column 1 implies that there is large significant effect of ownership rights on valuation when owners' cash-flow share is nil. But a cash-flow share of nil is meaningless. Better news for the published paper is that the interaction terms which are the authors' main focus clearly are robustly estimated.

Finally, in Table 11, we explore a specific set of results from Spilimbergo (2009), that the interaction of "students abroad" with "democracy in host country" has a negative effect on the Polity2 measure of democracy.[8] This is a panel-data analysis (country by time) with both time- and country-fixed effects which implies that the coefficients to

[7] The dateset used Castro, Clementi, and MacDonald (2004) is fairly small—45 observations—and some non-robustness must be expected. A fair discussion of the validity of their results would involve a much longer discussion.

[8] We choose this article because it is an example of panel data regression for which the data are easily available; however, the results we replicate are just one of a set of estimations in Spilimbergo's (2009) article so the discussion here should be seen as an example rather than an examination of the central message of Spilimbergo's paper.

the main terms, are determined by these variable after country and time means have been subtracted. We ask if the results are robust to potentially country-varying slopes to the main terms by removing country-specific averages before interacting. The first column shows the results reported in Spilimbergo (2009) while the second column replicates the analysis using the data posted by Spilimbergo on the web site of the *American Economic Review*—we need both, in order to ascertain that any deviation between our results and the results in the *American Economic Review* is not due the discrepancy between the posted data and the data actually used by Spilimbergo. The results are very similar between those columns except the R-square is much higher using the posted data. In column (3), we show the results using interactions that are demeaned country-by-country. The results are clearly not robust to this alternative specification—the coefficient to (non-interacted) "students abroad" becomes insignificant while the coefficient to the interaction changes from significantly negative to (nearly significantly) positive. Within the setting of our paper, it will take us too far afield to discuss in detail whether country-varying slopes in this setting is a reasonable alternative empirical specification for Spilimbergo's study, although it does not seem far fetched that growth of, say, democracy, varies across countries. Our main point is that, in general, in panel studies using data from heterogenous cross-sectional units it may be a reasonable alternative (unless rules out by theory) and it will often be reasonable to examine robustness against this alternative.

4 Conclusions

We provide practical advice regarding interpretation and robustness of models with interaction terms for econometric practitioners—in particular, we suggest some simple rules-of-thumb intended to minimize the risk of estimated interaction terms spuriously capturing other features of the data. The main tenet of our results is that researchers applying interaction terms should be very careful with specification and interpretation and not just put $X_1 X_2$ into a regression equation without considering robustness of results to functional form.

Appendix—Notes on data collection

Rajan and Zingales (1998):

The data is downloaded from Luigi Zingales' home page. The dependent variable is the *annual compounded growth rate* in real value added for each ISIC industry in each country for the period 1980–1990. *External dependence* (E) is the fraction of capital expenditures not financed with internal funds for firms in the United States in the same industry between 1980 and 1990. *Total capitalization* (T) is the ratio of (equity market capitalization plus domestic credit) to GDP. *Accounting standards* (A) is a country-level index developed by the Center for International Financial Analysis and Research ranking the amount of disclosure in annual company reports. I is *industry's share* of total value added in manufacturing in 1980 is from the United Nations Statistics. For more details on data sources, see Rajan and Zingales (1998).

Rajan and Zingales (2003):

We collected the data using the sources given in Rajan and Zingales (2003). The dependent variable, *number of companies to population*, is the ratio of the number of domestic companies whose equity is publicly traded in a domestic stock exchange to population in millions in 1993 (it is used as an indicator of the importance of equity markets). As a first source, stock exchange handbooks are used to count the number of companies and the Bulletin of the International Institute of Statistics is used as a second source. The countries in the sample are Australia, Austria, Belgium, Brazil, Canada, Denmark, France, Germany, India, Italy, Japan, the Netherlands, Norway, Russia, Sweden, Switzerland, the UK, and the United States.

GDP is Gross Domestic Product in 1913 obtained from International Historical Statistics (Mitchell, 1995). We could not find this series for Russia and we used figure 2 in Rajan and Zingales (2003) to interpolate the data. *Openness* (O) is the sum of exports and imports of goods in 1913 divided by GDP in 1913. Exports and imports are from the Statistical Yearbook of the League of Nations.[9] For Brazil and Russia, we could not find export and import data and we interpolated them from the averages of the variables in Rajan and Zingales (2003)'s Table 6.

Per capita industrialization (I) is the index of industrialization by country in 1913 as computed by Bairoch (1982). For more details about data sources, see Rajan and Zingales (2003).

Castro, Clementi, and MacDonald (2004):

We collected the data using the sources given in Castro, Clementi, and MacDonald (2004). The dependent variable is the average annual growth rate of real GDP per worker 1967-1996. *Real GDP per worker* is from the Penn World Tables, version 6.1. The set of countries corresponds to the 49 countries in La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1998) except we do not have data for Germany, Jordan, Venezuela, Switzerland, Zimbabwe, and Taiwan. Castro, Clementi, and MacDonald (2004) use four of the indicators of nvestor protection introduced by La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1998). The variable CR is an index aggregating different creditor rights in firm reorganization and liquidation upon default. The indicator antidirector rights (AR), and the dummy one share-one vote (OV), are two indices of shareholder rights geared towards measuring the ability of small shareholders to participate in decision making. Finally, the index rule of law (RL), proxies for the quality of law enforcement. These variables are described in more details in La Porta, Lopez-de-Silanes, Shleifer, and Vishny (1998. (1998).

RCT is a variable created to measure restrictions on capital transactions. First, a time-series dummy is constructed based on the IMF's Annual Report on Exchange Arrangements and Exchange Restrictions. The dummy variable takes the value of 1 for a given country in a given year if the IMF finds evidence of restrictions on payments on capital transactions for that country-year. Such restrictions include both taxes and

[9]See $http://www.library.northwestern.edu/govpub/collections/leaque/stat.html$.

quantity restrictions on the trade of foreign assets. Second, we compute RCT as the average of this dummy over the sample period to obtain a measure of the fraction of time each country imposed restrictions on international capital transactions.

Caprio, Laeven, and Levine (2007):

The exact data is used in Caprio, Laeven, and Levine (2007) and is downloaded from Ross Levine's home-page. It is a new database on bank ownership around the world and constructed by Caprio, Laeven, and Levine (2007). *Market-to-book* is the market to book value of each bank's equity of a bank from Bankscope database published in 2003.[10] In other words, it is the ratio of the market value of equity to the book value of equity. *Loan Growth* (LG) is each bank's average net loan growth during the last 3 years from Bankscope published in 2003.

Rights is an index of anti-director rights for the country from La Porta, Lopez-de-Silanes, Shleifer, and Vishny (2002). The range for the index is from zero to six formed by adding by adding the number of times each of the following conditions hold: (1) the country allows shareholders to mail their proxy vote, (2) shareholders are not required to deposit their shares prior to the General Shareholders' Meeting, (3) cumulative voting or proportional representation of minorities on the board of directors is allowed, (4) an oppressed minorities mechanism is in place, (5) the minimum percentage of share capital that entitles a shareholder to call for an Extraordinary Shareholders' Meeting is less than or equal to 10 percent (the sample median), or (6) when shareholders have preemptive rights that can only be waived by a shareholders meeting.

CF is the fraction of each bank's ultimate cash-flow rights held by the controlling owners. CF values are computed as the product of all the equity stakes along the control chain. The controlling shareholder may hold cash-flow rights directly (i.e., through shares registered in her name) and indirectly (i.e., through shares held by entities that, in

[10]Bankscope, maintained by Bureau van Dijk, contains financial and ownership information for about 4,000 major banks.

turn, she controls). If there is a control chain, then we use the products of the cash-flow rights along the chain. To compute the controlling shareholders total cash-flow rights we sum direct and all indirect cash-flow rights.[11] See Caprio, Laeven, and Levine (2007) for more details on data sources.

Spilimbergo (2009):

The exact data is used in Spilimbergo (2009) and is available from the *American Economic Review*'s web site. It is a unique panel data set of foreign students. The data forms an unbalanced panel comprising five year intervals between 1955 and 2000. The dependent variable, *Polity*2, is an index of democracy. *Students Abroad* (S) is the share of foreign students over population and *Democracy in host countries* (DH) is the average democracy index in host countries. See Spilimbergo (2009) for more details on data sources.

[11] Caprio, Laeven, and Levine (2007)'s calculations are based on Bankscope, Worldscope, the Bankers' Almanac, 20-F filings, and company web sites.

References

Allison, Paul D. (1977), "Testing For Interaction in Multiple Regression," *American Journal of Sociology* 83, 144–153.

Althauser, Robert (1971), "Multicollinearity and Non-Additive Regression Models," 453–472 in H. M. Blalock, Jr.(ed). *Causal Models in the Social Sciences,* Chicago: Aldine-Atherton.

Bairoch, Paul (1982), "International Industrialization levels from 1750 to 1980," *Journal of European Economic History* 11, 269–334.

Braumoeller, Bear F. (2004), "Hypothesis Testing and Multiplicative Interaction Terms," *International Organization* 58, 807–820.

Burnside, Craig and David Dollar (2000), "Aid, Policies, and Growth," *American Economic Review* 90, 847–868.

Caprio, Gerard, Luc Laeven, and Ross Levine (2007), "Governance and Bank Valuation," *Journal of Financial Intermediation* 16, 584–617.

Castro, Rui, Gian Luca Clementi and Glenn MacDonald (2004), "Investor Protection, Optimal Incentives, and Economic Growth," *Quarterly Journal of Economics* 119, 1131–1175.

Frisch, Ragnar and Frederick V. Waugh (1933), "Partial Time Regressions as Compared with Individual Trends," *Econometrica* 1, 387-401.

Heston, Alan, Robert Summers, and Bettina Aten (2002), Penn World Table Version 6.1 (Philadelphia, PA: Center for International Comparisons at the University of Pennsylvania).

Knack, Stephen and Philip Keefer (1995), "Institutions and Economic Performance: Cross-Country Tests Using Alternative Institutional Measures," *Economics and Policies* 7, 207–227.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny (1998), "Law and Finance," *Journal of Political Economy* 106, 1113–1155.

La Porta, Rafael, Florencio Lopez-de-Silanes, Andrei Shleifer, and Robert Vishny (2002), "Investor Protection and Corporate Valuation," *Journal of Finance* 57, 1147–1170.

Mitchell, Brian R. (1995), *International Historical Statistics*, Stockton Press, London.

Rajan, Raghuram G. and Luigi Zingales (1998), "Financial Dependence and Growth," *American Economic Review* 88, 559–589.

Rajan, Raghuram G. and Luigi Zingales (2003), "The Great Reversals: The Politics of Financial Development in the Twentieth Century," *Journal of Financial Economics* 69, 5–50.

Smith, Kent W. and M.S. Sasaki (1979), "Decreasing Multicollinearity: A Method for Models With Multiplicative Functions," *Sociological Methods and Research* 8, 35–56.

Spilimbergo, Antonio (2009), "Democracy and Foreign Education," *American Economic Review* 99, 528–543.

Table 1: **Simulation of Models**

Dependent Variable: Y

True model is $Y = 3\ X_1 + 5\ X_2 + 8\ X_1\, X_2 + \epsilon$

	(1)	(2)	(3)
X_1	10.989	2.999	10.997
	(19.202)	(4.72)	(24.51)
X_2	12.994	4.996	12.991
	(22.71)	(7.86)	(28.97)
$X_1\, X_2$	–	8.000	–
		(17.75)	
$(X_1 - \overline{X_1})\,(X_2 - \overline{X_2})$	–	–	8.000
			(17.75)
R^2	0.64	0.78	0.78

Notes: The true model is $Y = 3\ X_1 + 5\ X_2 + 8\ X_1\, X_2 + \epsilon$ where $X_1 = 1 + \epsilon_1$ and $X_2 = 1 + \epsilon_2$, $\epsilon_i \sim N(0, 1)$ for i=1, 2 (X_1 and X_2 are not correlated) and $\epsilon \sim N(0, 100)$. A constant is included but not reported. The sample size is 500 and the number of simulations is 20000. The averages of the estimated t-statistics are shown in parentheses.

Table 2: **Simulation of Models: Misspecified Model**

Dependent Variable: Y

True model is $Y = X_1 + X_1^2 + \epsilon$

	(1)	(2)	(3)
X_1	1.000	3.001	1.000
	(12.84)	(36.77)	(6.98)
X_2	–	0.000	−0.000
		(0.00)	(−0.00)
X_1^2	1.000	–	1.000
	(31.38)		(15.57)
X_2^2	–	–	0.000
			(0.00)
$(X_1 - \overline{X_1})(X_2 - \overline{X_2})$	–	0.666	0.000
		(19.88)	(0.00)
R^2	0.92	0.86	0.92

Notes: True model is $Y = X_1 + X_1^2 + \epsilon$ where $X_1 = 1 + \epsilon_1$ and $X_2 = 1 + X_1 + \epsilon_2$, $\epsilon_i \sim N(0,1)$ for all i (X_1 and X_2 are correlated). A constant is included but not reported. The sample size is 500 and the number of simulations is 20000. The averages of the estimated t-statistics are shown in parentheses.

Table 3: **Simulation of Models: PANEL**

Dependent Variable: Y

True model is $Y_{it} = \alpha_i + X_{1it} + \xi_i\ X_{2it} + \epsilon_{it}$

	(1)	(2)
X_1	1.000	1.000
	(26.57)	(25.80)
X_2	1.500	1.500
	(56.36)	(54.72)
$(X_1 - \overline{X}_{1..})\,(X_2 - \overline{X}_{2..})$	-0.152	–
	(-11.01)	
$(X_1 - \overline{X}_{1i.})\,(X_2 - \overline{X}_{2i.})$	–	-0.000
		(-0.02)
R^2	0.86	0.85

Notes: True model is $Y_{it} = \alpha_i + X_{1it} + \xi_i\ X_{2it} + \epsilon_{it}$ where $X_{11t} = 1 + \epsilon_{1t}$ and $X_{21t} = 1 + X_{11t} + \epsilon_{2t}$ for the first country, $X_{12t} = 1/4 + \epsilon_{3t}$ and $X_{22t} = 1 + X_{12t} + \epsilon_{4t}$ for the second country where $\epsilon_{it} \sim N(0,1)$ for all i. X_1 and X_2 are correlated for each countries. We let $\xi_1 = 1$ and $\xi_2 = 2$. Fixed effects are included in the regressions but not reported. We have $i = 1, 2$ and $t = 1, ..., 500$. The number of simulations is 20000. The averages of the estimated t-statistics are shown in parentheses.

Table 4: **Simulation of Models: Frisch-Waugh — A**

Dependent Variable: Y

True model is $Y = 3\ X_1 + 5\ X_2 + 8\ X_1\, X_2 + \epsilon$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X_1	18.993 (17.65)	18.999 (29.83)	18.999 (18.30)	19.004 (19.96)	18.993 (19.94)	18.999 (29.83)	18.999 (29.83)	18.999 (29.83)
X_2	13.006 (17.10)	13.002 (28.88)	13.001 (17.72)	12.998 (19.31)	13.004 (19.31)	13.002 (28.88)	13.002 (28.88)	13.002 (28.88)
$(X_1 - \overline{X_1})\,(X_2 - \overline{X_2})$	–	7.994 (10.17)	–	–	–	–	–	–
$(X_1 - \overline{X_1})^2$	–	0.009 (0.01)	–	–	–	5.338 (14.40)	0.009 (0.01)	8.006 (17.56)
$(X_2 - \overline{X_2})^2$	–	0.000 (−0.00)	–	–	–	2.664 (14.38)	3.997 (17.54)	0.000 (−0.00)
$X_1^{\psi}\, X_2^{\psi}$	–	–	5.477 (4.00)	–	–	5.331 (10.17)	–	–
$X_1^{\psi}\,(X_2 - \overline{X_2})$	–	–	–	8.021 (7.42)	–	–	7.994 (10.17)	–
$(X_1 - \overline{X_1})\, X_2^{\psi}$	–	–	–	–	8.045 (7.45)	–	–	7.994 (10.17)
R^2	0.81	0.93	0.82	0.85	0.85	0.93	0.93	0.93

Notes: True model is $Y = 3\ X_1 + 5\ X_2 + 8\ X_1\, X_2 + \epsilon$ or, equivalently, $Y = 19\ X_1 + 13\ X_2 + 8\ (X_1 - \overline{X_1})\,(X_2 - \overline{X_2}) + \epsilon$ where $X_1 = 1 + \epsilon_1$ and $X_2 = 1 + X_1 + \epsilon_2$ where $\epsilon_i \sim N(0,1)$ for all i=1,2 (X_1 and X_2 are correlated) and $\epsilon \sim N(0,100)$. For columns (4) – (7); $X_1^{\psi} = M_2 = [I - P_{[constant, X_2]}]\, X_1$, $X_2^{\psi} = [I - P_{[constant, X_1]}]\, X_2$. A constant is included but not reported. The sample size is 500 and the number of simulations is 20000. The averages of the estimated t-statistics are shown in parentheses.

Table 5: **Simulation of Models: Frisch-Waugh — B**

Dependent Variable: Y

True model is $Y = 3\ X_1 + 5\ X_2 + 8\ X_1\ \epsilon_2 + \epsilon$

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
X_1	−4.998 (−6.18)	−4.999 (−6.61)	−5.001 (−6.61)	−4.998 (−6.17)	−4.997 (−7.87)	−4.996 (−7.85)	−4.996 (−7.85)	−4.996 (−7.85)
X_2	12.998 (22.71)	12.996 (24.32)	13.001 (24.28)	12.999 (22.69)	12.994 (28.92)	12.994 (28.85)	12.994 (28.85)	12.994 (28.85)
$(X_1 - \overline{X_1})\,(X_2 - \overline{X_2})$	–	2.667 (8.59)	–	–	–	–	–	–
$(X_1 - \overline{X_1})^2$	–	–	–	–	–	−2.671 (−7.23)	−8.005 (−12.46)	−0.001 (−0.04)
$(X_2 - \overline{X_2})^2$	–	–	–	–	–	2.668 (14.39)	4.004 (17.54)	−0.001 (−0.00)
$X_1^{\psi}\ X_2^{\psi}$	–	–	5.340 (8.57)	–	–	5.338 (10.17)	–	–
$X_1^{\psi}\,(X_2 - \overline{X_2})$	–	–	–	−0.015 (−0.03)	–	–	8.005 (10.17)	–
$(X_1 - \overline{X_1})\,X_2^{\psi}$	–	–	–	–	8.003 (17.69)	–	–	8.005 (10.17)
R^2	0.59	0.64	0.64	0.59	0.75	0.75	0.75	0.75

Notes: The true model is $Y = 3\ X_1 + 5\ X_2 + 8\ X_1\ \epsilon_2 + \epsilon$ where $X_1 = 1 + \epsilon_1$ and $X_2 = 1 + X_1 + \epsilon_2$ where $\epsilon_i \sim N(0,1)$ for all i=1,2 (X_1 and X_2 are correlated) and $\epsilon \sim N(0,100)$. For columns (4) – (7); $X_1^{\psi} = [I - P_{[constant,X_2]}]\,X_1,\ X_2^{\psi} = [I - P_{[constant,X_1]}]\,X_2$. A constant is included but not reported. The sample size is 500 and the number of simulations is 20000. The average of the estimated t-statistics are shown in parentheses.

Table 6: **Simulation of Models: Frisch-Waugh — Misspecified Model**

Dependent Variable: Y

True model is $Y = X_1 + X_1^2 + \epsilon$

	(1)	(2)	(3)	(4)	(5)
X_1	1.000	3.000	3.001	3.001	3.001
	(12.84)	(36.79)	(27.47)	(27.47)	(27.47)
X_2	–	0.000	0.001	0.001	0.001
		(0.00)	(0.01)	(0.01)	(0.01)
X_1^2	1.000	–	–	–	–
	(31.40)				
$(X_1 - \overline{X_1})\,(X_2 - \overline{X_2})$	–	0.666	–	–	–
		(19.88)			
$X_1^{\psi}\,X_2^{\psi}$	–	–	0.005	–	–
			(0.05)		
$X_1^{\psi}\,(X_2 - \overline{X_2})$	–	–	–	1.000	–
				(15.70)	
$(X_1 - \overline{X_1})\,X_2^{\psi}$	–	–	–	–	0.001
					(0.01)
R^2	0.92	0.86	0.75	0.83	0.75

Notes: True model is $Y = X_1 + X_1^2 + \epsilon$ where $X_1 = 1 + \epsilon_1$ and $X_2 = 2 + X_1 + \epsilon_2$ where $\epsilon_i \sim N(0,1)$ for all i (X_1 and X_2 are correlated). For columns (1) – (2); $X_1^{\psi} = M_2 = [I - P_{[constant, X_2]}]\,X_1$, $X_2^{\psi} = [I - P_{[constant, X_1]}]\,X_2$. A constant is included but not reported. The sample size is 500 and the number of simulations is 20000. The average of the estimated t-statistics in parentheses.

Table 7: **Replication of Rajan and Zingales (1998): Table 4 (column 5)**

Dependent Variable: Annual compounded growth rate

	(1)†	(2)
I	-4.33	-4.33
	(-3.20)	(-3.20)
$E\,T$	0.12	–
	(0.82)	
$E\,A$	1.33	–
	(3.74)	
$(E-\overline{E})\,T^{\psi}$	–	0.38
		(2.40)
$(E-\overline{E})\,A^{\psi}$	–	1.73
		(4.38)
R^2	0.42	0.42

Notes: The column marked with † replicates Rajan and Zingales (1998). The dependent variable is the annual compounded growth rate in real value added for each ISIC industry in each country for the period 1980–1990. External dependence, E, is the fraction of capital expenditures not financed with internal funds for U.S. firms in the same industry between 1980–1990. For interaction terms, external dependence is multiplied by financial development variables; total capitalization to GDP ratio (T) and accounting standards in a country in 1990 (A). T is the ratio of the sum of equity market capitalization and domestic credit to GDP (It varies by country). I is industry's share of total value added in manufacturing in 1980. The sample size is 1042 for all of the regressions. All regressions include a constant, country- and industry-fixed effects but their coefficients are not reported. All coefficients are multiplied by 10. t-statistics in parentheses. The new variables which are created according to the Frisch-Waugh theorem: $E^{\psi} = (I - P_{[constant,I,T,A,countrydummies]})\,E$, $T^{\psi} = (I - P_{[constant,I,E,A]})\,T$, $A^{\psi} = (I - P_{[constant,I,E,T]})\,A$.

Table 8: **Replication of Rajan and Zingales (2003): Table 7 (Panel B)**

Dependent Variable: Number of companies/million population

	(1)†	(2)†	(3)	(4)	(5)	(6)	(7)
I	238.46	−212.00	362.44	318.03	354.22	370.00	347.27
	(1.76)	(−1.37)	(3.49)	(0.72)	(3.34)	(3.46)	(3.37)
O	35.36	−0.91	44.17	69.00	41.05	40.65	44.59
	(3.86)	(−0.08)	(6.26)	(2.30)	(5.86)	(5.85)	(6.27)
I^2	–	–	–	0.07	–	–	–
				(0.02)			
O^2	–	–	–	−10.58	–	–	–
				(−0.87)			
$I\,O$	–	919.95	–	–	–	–	–
		(3.79)					
$(I-\overline{I})\,(O-\overline{O})$	–	–	919.95	743.35	–	–	–
			(3.79)	(2.27)			
$I^\psi\,O^\psi$	–	–	–	–	957.24	–	–
					(3.61)		
$I^\psi\,(O-\overline{O})$	–	–	–	–	–	950.65	–
						(3.64)	
$(I-\overline{I})\,O^\psi$	–	–	–	–	–	–	929.71
							(3.76)
R^2	0.54	0.77	0.77	0.79	0.76	0.76	0.77

Notes: The column marked with † replicates Rajan and Zingales (2003). The dependent variable is the number of listed companies per million of population in 1913. Per capita industrialization (I) is the index of industrialization for that country in 1913. Openness (O) is the sum of exports and imports of goods divided by GDP in 1913. Coefficient estimates for per capita industrialization and its interaction with openness are multiplied by 1000. A constant is included in all regressions but not reported. t-statistics in parentheses. The new variables which are created according to the Frisch-Waugh theorem are $I^\psi = M_2 = [I - P_{[\beta_0,O]}]\,I$ and $O^\psi = [I - P_{[\beta_0,I]}]\,O$.

Table 9: **Replication of Castro, Clementi, and MacDonald (2004): Table 1**

Dependent Variable: Average annual growth rate of real GDP per worker 1967-1996

	(1)†	(2)	(3)	(4)	
$OV * RCT$	9.10	−9.94	3.36	–	–
	(0.60)	(−0.57)	(0.19)		
$AR * RCT$	−5.30	−7.29	−5.78	–	–
	(−1.42)	(−1.64)	(−1.35)		
$CR * RCT$	−10.29	−11.14	−3.73	–	–
	(−2.17)	(−2.30)	(−0.64)		
$RL * RCT$	−1.01	1.22	2.51	–	–
	(−0.34)	(0.35)	(0.73)		
$LRGDPW67^2$	–	–	−9.25	–	–
			(−2.02)		
$OV^{\psi}\,(RCT - \overline{RCT})$	–	–	–	15.57	–
				(1.02)	
$AR^{\psi}\,(RCT - \overline{RCT})$	–	–	–	−6.15	–
				(−1.55)	
$CR^{\psi}\,(RCT - \overline{RCT})$	–	–	–	−9.38	–
				(−1.32)	
$RL^{\psi}\,(RCT - \overline{RCT})$	–	–	–	−12.62	–
				(−2.06)	
$(OV - \overline{OV})\,RCT^{\psi}$	–	–	–	–	7.14
					(0.40)
$(AR - \overline{AR})\,RCT^{\psi}$	–	–	–	–	−6.57
					(−1.52)
$(CR - \overline{CR})\,RCT^{\psi}$	–	–	–	–	−14.30
					(−2.60)
$(RL - \overline{RL})\,RCT^{\psi}$	–	–	–	–	−2.91
					(−0.91)
Quadratic terms included but not shown	N	Y	Y	N	N
R^2	0.50	0.59	0.64	0.50	0.53

Notes: The column marked with † replicates Castro, Clementi, and MacDonald (2004). RCT measures restrictions on capital transactions. CR is an index of creditor rights. AR is an indicator of antidirector rights and OV is a dummy for one-share one-vote, OV. $LRGDPW67$ is the natural logarithm of real gross domestic product per worker in 1967. RL is an index for rule of law. The coefficients are multiplied by 1000. Main terms of these variables are included in all regressions but suppressed. Sample size is 43. A constant is included but not reported. t-statistics in parentheses. The new variables which are created according to the Frisch-Waugh theorem are: $RCT^{\psi} = (I - P_{[constant,LRGDPW67,OV,AR,CR,RL]})\,RCT$, $OV^{\psi} = (I - P_{[constant,LRGDPW67,RCT,AR,CR,RL]})\,OV$, $AR^{\psi} = (I - P_{[constant,LRGDPW67,OV,RCT,CR,RL]})\,AR$, $CR^{\psi} = (I - P_{[constant,LRGDPW67,OV,AR,RCT,RL]})\,CR$, and $RL^{\psi} = (I - P_{[constant,LRGDPW67,OV,AR,CR,RCT]})\,RL$.

Table 10: **Replication of Caprio, Laeven, and Levine (2007): Table 5 (Column 1)**

Dependent Variable: Market-to-book value

	(1)†	(2)	(3)	(4)
Loan Growth	0.27	0.64	0.19	0.19
	(0.76)	(1.42)	(0.54)	(0.53)
Rights	0.31	−0.22	0.14	0.07
	(5.75)	(−1.10)	(3.43)	(1.81)
CF	2.27	−1.57	−0.57	−0.44
	(4.33)	(−2.98)	(−3.23)	(−2.35)
$Loan\ Growth^{\ 2}$	0.27	−1.17	–	–
	(0.76)	(−1.44)		
$Rights^{\ 2}$	–	0.05	–	–
		(1.51)		
$CF^{\ 2}$	–	1.34	–	–
		(2.03)		
CF Rights	−0.89	–	–	–
	(−5.78)			
$(CF - \overline{CF})\,(Rights - \overline{Rights})$	–	−0.82	–	–
		(−5.11)		
$CF^{\ \psi}\,(Rights - \overline{Rights})$	–	–	−0.91	–
			(−6.14)	
$(CF - \overline{CF})\ Rights^{\ \psi}$	–	–	–	−0.39
				(−4.32)
R^2	0.20	0.23	0.22	0.15

Notes: The column marked with † replicates Caprio, Laeven, and Levine (2007). A constant is included in all regressions but not reported. *Market-to-book* is the market to book value of the bank's equity of a bank. *Loan Growth* (*LG*) is the bank's average net loan growth during the last 3 years. *Rights* is an index of anti-director rights for the country. *CF* is the fraction of the bank's ultimate cash-flow rights held by the controlling owners. Sample size is 213. *t*-statistics in parentheses. The new variables which are created according to the Frisch-Waugh theorem are: $CF^{\psi} = (I - P_{[c,LG,Rights]})\,CF$ and $Rights^{\psi} = (I - P_{[c,LG,CF]})\,Rights$.

Table 11: **Replication of Spilimbergo (2009): Table 2a (Column 2)**

Dependent Variable: *Polity*2 index of democracy

	$(1)^{\dagger}$	(2)	(3)
$Democracy_{t-5}$	0.45	0.44	0.44
	(9.61)	(8.46)	(8.44)
$Students\ Abroad_{t-5}\,(S)$	24.23	24.23	−1.82
	(2.81)	(2.55)	(−0.39)
$Democracy\ in\ host\ countries_{t-5}\,(DH)$	0.12	0.12	0.10
	(2.23)	(2.23)	(1.84)
$S_{t-5}\ DH_{t-5}$	−33.71	−33.31	–
	(−2.71)	(−2.47)	
$(S_{t-5} - \overline{S_{t-5}}_{i.})\,(DH_{t-5} - \overline{DH_{t-5}}_{i.})$	–	–	56.44
			(1.73)
Time effects	Yes	Yes	Yes
Country effects	Yes	Yes	Yes
Observations	1107	1121	1121
R^2	0.41	0.82	0.82

Notes: The column marked with † replicates Spilimbergo (2009). The data forms an unbalanced panel comprising five year intervals between 1955 and 2000. The dependent variable, *Polity*2, is the composite Polity II democracy index from the Polity IV data set. *Students Abroad* (S) is the share of foreign students over population and *Democracy in host countries* (DH) is the average democracy index in host countries. t-statistics in parentheses.