

Interaction Effects in Econometrics

Hatice Ozer-Balli *

Massey University

Bent E. Sørensen †

University of Houston and CEPR

25 June 2010

Abstract

We provide practical advice for applied economists regarding specification and interpretation of linear regression models with interaction terms.

JEL classification: C12, C13

Keywords: Non-Linear Regression, Interaction Terms.

*School of Economics and Finance, Massey University, New Zealand, e-mail: h.ozerballi@massey.ac.nz, tel:+64 63505799 ext. 2666.

†Department of Economics, University of Houston, TX, e-mail: bent.sorensen@mail.uh.edu, tel: 7137433841, fax: 7137433798

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.

In Section 2, we discuss some practical issues related to the specification of regressions with interaction effects and make recommendations for practitioners. In Section 3, we illustrate our recommendations with Monte Carlo simulations and, in Section 4, we revisit some prominent applied papers where interaction effects figure prominently, including Rajan and Zingales (1998), and examine if the published results are robust. Section 5 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) and instrumental variable (IV) 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 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:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon, \quad (1)$$

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 grows 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:

1. In a regression with interaction terms, the main terms should always be included. 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.)¹
2. 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.

¹Some authors have referred to this as a multicollinearity problem. Althauser (1971) show 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.

In applied papers, the non-interacted regression

$$Y = \lambda_0 + \lambda_1 X_1 + \lambda_2 X_2 + v, \quad (2)$$

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 $X_2 = \bar{X}_2$ (the mean value of X_2).² The estimated β_1 -coefficient in (1) is typically very close to $\hat{\lambda}_1 - \hat{\beta}_3 \bar{X}_2$.

3. Estimating the interacted regression in the form

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

results in the exact same fit as equation (1) and the exact same coefficient $\hat{\beta}_3$. $\hat{\beta}_1$ 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 $X_2 = \bar{X}_2$. If a researcher reports results from (2), and wants to keep the interpretation of the coefficient to main terms similar, is usually preferable to report results of the regression (3) with demeaned interaction terms.³

4. 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 $X_1 X_2$. Alternatively, one can regress X_2 on Z and obtain \hat{X}_2 and use $X_1 \hat{X}_2$ for the interaction term and obtain a consistent estimate of β_3 .

²Some 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), Althausser (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.

³Because $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 (X_1 - \bar{X}_1) (X_2 - \bar{X}_2) = (\beta_0 + \beta_3 \bar{X}_1 \bar{X}_2) + (\beta_1 - \beta_3 \bar{X}_2) X_1 + (\beta_2 - \beta_3 \bar{X}_1) X_2 + \beta_3 X_1 X_2$, we get the exact same fit, with the changes in the estimated parameters given from the correspondence between the left- and right-hand side of this equality. E.g., $\hat{\lambda}_0$ will be equal to $\hat{\beta}_0 + \beta_3 \bar{X}_1 \bar{X}_2$.

2.1 Robustness to misspecification

Often a researcher wants to test whether $Y = f(X_1, X_2)$ and chose a linear specification such as (2) for convenience. A more adequate specification may be a second order expansion

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

(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.

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 $\text{corr}(X_1, X_2) \neq 0$, the coefficient β_3 in the interacted regression (1) may be spuriously significant. For example, if $\text{corr}(X_1, X_2) > 0$ 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 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 $\alpha \delta$.

2.2 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 \bar{X}_i . (i.e., $\frac{1}{T}\sum_{t=1}^T X_{it}$), the average across cross-sectional units at period t by $\bar{X}_{.t}$, and the mean across all observations by $\bar{X}_{..}$.

Consider the panel data regression

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

where μ_i and ν_t 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,

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

then, if the mean of X_1 varies by country and the covariance of \bar{X}_{1i} and β_{2i} is non-zero, the covariance of $(X_{1it} - \bar{X}_{1..}) (X_{2it} - \bar{X}_{2..})$ 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

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

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.)

3 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 which we restate here for convenience.

3.1 Frisch-Waugh theorem

In the multiple linear regression model, it is common that the interest centers on a subset of the full variables. Consider an equation

$$Y = X_1\beta_1 + X_2\beta_2 + \epsilon, \tag{7}$$

where β_1 is $k_1 \times 1$, β_2 is $k_2 \times 1$. Let's say our concern is just to find β_1 coefficients.

Theorem (Frisch-Waugh):⁴ *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 identical to the estimate from the formula*

$$\hat{\beta}_1 = (X_1^{\psi'} X_1^{\psi})^{-1} X_1^{\psi'} Y^{\psi},$$

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 .

⁴See Frisch and Waugh (1933) for details.

3.2 Using the Frisch-Waugh theorem to hedge against a spurious interaction term

Consider equation (1). If we want to find 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 our concern is how the variable X_1 , cleaned of any other regressors, affects the impact of X_2 on Y , we suggest running the following regression model:

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

where $X_1^\psi = M_2 X_1$, $M_2 = [I - P_{\beta_0, X_2}]$ where M_2 is the residual maker (from regressing X_1 on a constant and X_2).

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 are significant it makes it less likely that the interaction is spurious.

Case 2: if one wants to ascertain that the interaction of X_1 and X_2 captures no other regressors the safest strategy is to run the following regression model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1^\psi X_2^\psi + \epsilon, \quad (9)$$

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 a residual maker; regressing X_2 on a constant and X_1 and M_2 is the residual maker; regressing X_1 on a constant and X_2).

In general, there is no way of obtaining a consistent estimate of the interaction effect without knowing the exact generating process. In applications, interaction effects are

however often intuitively motivated and we will illustrate in the Monte Carlo section how different generating processes will affect inference.

4 Monte Carlo Simulations

In this section we illustrate how our recommendations allow for robust inference using Monte Carlo simulations. We run 10,000 simulations of sample size 200.

4.1 Interpretation of the main terms

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 the 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.

4.2 IV estimation

Next we consider a model with an interaction effect where one of the independent variables is endogenous. We generate the true model as $Y = X_1 + X_2 + X_1 X_2 + \epsilon_2$, where $X_1 = \epsilon_1$, $X_2 = 0.5 Z_1 + 0.5 X_1 + \epsilon_2$. X_2 is an endogenous variable which is correlated with X_1 . A valid instrument for X_2 is $Z_1 = \epsilon_3$.

In Table 2, we show OLS and IV regressions, starting with OLS-estimates of model (1). The coefficients are, as expected, severely biased. In column (2), X_2 is instrumented by Z_1 . Using the fitted value, \hat{X}_2 , from a regression of X_2 on *constant*, X_1 , Z_1 , Y is regressed on $X=[c, X_1, \hat{X}_2, X_1 \hat{X}_2]$ and we see that the parameters are all estimated without bias. In column (3), X_2 and $X_1 X_2$ (the interaction term) are instrumented by Z_1 and Z_3 , where $Z_3 = Z_1 X_1$. The regression delivers point estimates similar to those of column (2), but this regression uses the exogenous X_1 less directly in the interaction so this estimator is less efficient. In column (4), the regression do not make use of the exogenous variable X_1 to construct an instrument for the interaction term instead Z_2 is used where $Z_2 = Z_1 + \epsilon_4$ and there is a large loss of efficiency. In general, an IV-estimator is more efficient the higher the correlation of the instrument with the endogenous variable and in most applications one can expect $X_1 \hat{X}_2$ to have the highest correlation with $X_1 X_2$.

4.3 Non-linear terms in the regression

In Table 3, the true model doesn't 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, (3), and we see 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.

4.4 Panel data with varying slopes

We consider a panel data regression with two “countries” $i = 1, 2$ for $T = 200$ “years.” The true model have the slope for X_2 varying across countries: $Y_{it} = \alpha_i + X_{1it} + \xi_i X_{2it} + \epsilon_{it}$.⁵

In Table 4, 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 $\partial Y_{it} / \partial X_2 = \beta_2 + \beta_3 X_{1it}$. 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.

4.5 Frisch-Waugh orthogonalization

In Table 5, we simulate a model with an interaction term and correlated regressors and estimate various specifications as suggested above. Frisch-Waugh orthogonalization gives biased coefficients in columns (4) and (9), but the interaction remains significant. The most interesting column is column (9), which is technically biased with significant quadratic terms and too low an estimate of the interaction.⁶ However, the qualitative

⁵We 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_1 to vary by country by setting $\xi_1 = 1$ and $\xi_2 = 2$.

⁶The significant quadratic terms reflect that the interaction terms is correlated with quadratic terms when X_1 and X_2 are correlated.

message that researcher may draw from this may not be misleading and only a researcher with a very sharp (structural) model would draw misleading conclusions. While this is vague, a reading of the articles we replicate below will give examples of the not-so-sharp hypotheses typically examined using interaction terms.

Table 6 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 . In the example here, the regressions where X_2 is Frisch-Waugh orthogonalized deliver consistent estimates while the regular interaction, still significant, do not.

Table 7 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_1 or both of the independent variables have been orthogonalized.

5 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 8 we replicate the highly influential paper of Rajan and Zingales (1998). 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 sum of total equity market capitalization and 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 strengthens the size and sig-

nificance 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 9, we see that the t -statistic on the main terms are very much affected by the interaction terms not being centered. We also see that the impact of the interaction term is robust to including quadratic terms in the main variables.

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 where the latter effect is captured by interaction terms. Table 10 replicates Castro, Clementi, and MacDonald (2004)’s Table 1 and we see that the estimates of some, but not all, main terms are affected by the interaction terms not being centered; however, the general message, that higher growth correlates with better property rights, seems robust. Including quadratic terms in the property rights measures seem to strengthen the authors’ main result of negative interactions (although the inclusion of a quadratic term in GDP weakens it). If we use Frisch-Waugh residuals for either the creditor rights measures or 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 appear very robust to the kind of robustness checks that we recommend.

Caprio, Laeven, and Levine (2007) examine if bank valuations (relative to book values) are higher where owners have stronger right (*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 11 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, give a 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 term clearly is robustly estimated.

Easterly, Levine, and Roodman (2004) examine whether foreign aid (Aid) is more effective in countries with good policy ($Policy$). Their results indicate that this result is fragile to the sample, but here we focus on whether it is sensitive to the inclusion of quadratic terms etc. The first column of Table 12 replicates Easterly, Levine, and Roodman (2004)’s Table 1 (column 1) while the following columns add quadratic term, Frisch-Waugh orthogonalization, etc. The main focus is the interaction between aid and policy. Including quadratic terms strengthens the significance of the interaction of interest while the interaction becomes insignificant—with a point estimate equal to that of column (1)—when the Frisch-Waugh residual is used for aid. Using the Frisch-Waugh residual for policy strengthens the significance of the aid-policy interaction. Our permutations all make the point estimates for the main term in policy more significant and makes the main term for aid near-significant. Overall, for this particular sample, our battery of specification variations do not cast significant doubt on the notion that aid is more effective in countries with good policy.

Finally, in Table 13, we explore the results in Spilimbergo (2009) that the interaction of “students abroad” with “democracy in host country” has a negative effect on the Polity2 measure of democracy. This is a panel-data analysis (country by time) and we ask if the results are robust to potentially country-varying slopes to the main terms but

removing country-specific averages before interacting. The results are clearly not robust to this alternative specification.

6 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 and implications for the interpretation of results.

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 the U.S. firms in the same industry between years 1980 and 1990. *Total capitalization (T)* is the ratio of the sum of equity market capitalization and domestic credit to GDP. *Accounting standards (A)* is an index developed by the Center for International Financial Analysis and Research ranking the amount of disclosure in annual company reports in each country. *I* is *industry's share* of total value added in manufacturing in 1980 obtained from 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 the country's 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). For Russia, we couldn't find the data and we used figure 2 from 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. Both export and import

are obtained from the Statistical Yearbook of the League of Nations.⁷ For Brazil and Russia we couldn't find export and import data and we interpolated them looking at 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. *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. The time span of the sample is from 1967 to 1996. Investor protection is measured using indicators introduced in La Porta et al. (1998). Castro, Clementi, and MacDonald (2004) focus on four of these indicators. 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 et al. (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

⁷See <http://www.library.northwestern.edu/govpub/collections/league/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 et al. (2007). *Market-to-book* is the market to book value of each bank's equity of a bank from Bankscope database published in 2003.⁸ 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 et al. (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

⁸Bankscope, 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.⁹ See Caprio et al. (2007) for more details on data sources.

Easterly, Levine, and Roodman (2004):

The exact data used in Easterly, Levine, and Roodman (2004) and is downloaded from Ross Levine's home-page. The years covered are 1970-1997. *Aid* is the amount of international aid provided to a country. It is calculated by dividing the development assistance to real GDP. *Policy (P)*, an index of the quality of the policy environment, is a regression-weighted average of macroeconomic policies described in Burnside and Dollar (2000). It is constructed from measures of budget balance, inflation, and the Sachs-Warner openness index.

LGDP is logarithm of initial GDP per capita. *Ethnic (Et)* is ethnic fractionalization from Easterly and Levine (1997). It measures the probability that two randomly selected individuals in a country belong to different ethnolinguistic groups. *Assassinations (Ass)* is the rate of political assassinations per million population. *Sub-Saharan Africa (SSA)* and *Fast – growing E. Asia (FEA)* are regional dummy variables. *Institutional Quality (IQ)* is an index of institutional quality from Stephen Knack and Philip Keefer (1995). *M2/GDP lagged* is a measure of financial depth. For more details about data sources, see Easterly et al. (2004).

Spilimbergo (2009):

The exact data is used in Spilimbergo (2009) and it is available at the *American Economic Review's* web-page. It is a unique panel data set on foreign students. The data forms an unbalanced panel comprised at five year intervals between 1955 and 2000. The

⁹Caprio, Laeven, and Levine (2007)'s calculations are based on Bankscope, Worldscope, the Bankers' Almanac, 20-F filings, and company websites.

dependent variable, *Polity2*, is a measure of democracy index. *Polity2* is the composite Polity II index from the Polity IV data set which is the difference between the Polity's democracy and autocracy indices. *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 sources of data.

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.
- Easterly, William, and Ross Levine (1997), "Africa's Growth Tragedy: Policies and Ethnic Divisions," *Quarterly Journal of Economics* 112, 1203-1250.
- Easterly, William, Ross Levine, and David Roodman (2004), "Aid, Policies, and Growth: Comment," *American Economic Review* 94, 774–780.
- 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 Politics* 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	11.001 (12.17)	2.994 (2.96)	10.999 (15.44)
X_2	13.008 (14.39)	4.991 (4.93)	12.998 (18.25)
$X_1 X_2$	–	8.012 (11.12)	–
$(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$	–	–	8.012 (11.12)
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 200 and the number of simulations is 10000. The averages of the estimated t -statistics are shown in parentheses.

Table 2: **Simulation of Models: IV— X_2 is Endogeneous**

Dependent Variable: Y

True model is $Y = X_1 + X_2 + X_1 X_2 + \epsilon_2$

	OLS	IV	IV	IV
X_1	0.601 (17.17)	1.012 (4.44)	1.028 (4.08)	0.906 (2.35)
X_2	1.800 (63.04)	0.979 (3.16)	0.951 (2.82)	1.026 (1.85)
$X_1 X_2$	1.000 (40.93)	1.001 (5.42)	0.994 (2.88)	0.803 (0.43)
R^2	0.98	0.40	0.34	0.31

Notes: The true model is $Y = X_1 + X_2 + X_1 X_2 + \epsilon_2$ where $X_1 = \epsilon_1$, $X_2 = 0.5 Z_1 + 0.5 X_1 + \epsilon_2$. X_2 is an endogenous variable and it is correlated with X_1 . A constant is included but not reported. The sample size is 200 and the number of simulations is 10000. The averages of the estimated t -statistics are shown in parentheses. The instruments used in columns (2)-(4) includes: $Z_1 = \epsilon_3$, $Z_2 = Z_1 + \epsilon_4$, $Z_3 = Z_1 X_1$ where $\epsilon_i \sim N(0, 1)$ for all i .

In column (2), X_2 is instrumented by Z_1 . Using the fitted value of X_2 , \hat{X}_2 , Y is regressed on $X = [c X_1 \hat{X}_2 X_1 \hat{X}_2]$.

In column (3), both X_2 and $X_1 X_2$ (the interaction term) are instrumented by Z_1 and Z_3 . Using the fitted values of X_2 and interaction terms, \hat{X}_2 and $(X_1 \hat{X}_2)$, Y is regressed on $X = [c X_1 \hat{X}_2 (X_1 \hat{X}_2)]$.

In column (4), both X_2 and $X_1 X_2$ (the interaction term) are instrumented by Z_1 and Z_2 . Using the fitted values of X_2 and interaction terms, \hat{X}_2 and $(X_1 \hat{X}_2)$, Y is regressed on $X = [c X_1 \hat{X}_2 (X_1 \hat{X}_2)]$.

Table 3: **Simulation of Models: Misspecified Model**

Dependent Variable: Y

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

	(1)	(2)	(3)
X_1	1.001 (8.09)	3.000 (23.31)	1.005 (4.38)
X_2	–	0.001 (0.01)	0.001 (0.01)
X_1^2	0.999 (19.64)	–	0.997 (9.64)
X_2^2	–	–	–0.000 (–0.01)
$(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$	–	0.665 (12.48)	0.001 (0.02)
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 200 and the number of simulations is 10000. The averages of the estimated t -statistics are shown in parentheses.

Table 4: **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	1.001 (16.80)	1.001 (16.32)
X_2	1.499 (35.59)	1.499 (34.55)
$(X_1 - \bar{X}_{1..})(X_2 - \bar{X}_{2..})$	-0.153 (-6.96)	-
$(X_1 - \bar{X}_{1i.})(X_2 - \bar{X}_{2i.})$	-	-0.000 (-0.01)
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 $\xi_2 = 2$. We estimate the panel regression (5) but do not report the estimated fixed effects. We have $i = 1, 2$ and $t = 1, \dots, 200$. The number of simulations is 10000. The averages of the estimated t -statistics are in parentheses.

Table 5: **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)	(9)
X_1	18.988 (11.22)	18.997 (18.88)	18.998 (18.67)	18.983 (11.66)	18.989 (12.65)	18.981 (12.66)	18.998 (18.67)	18.998 (18.67)	18.998 (18.67)
X_2	13.000 (10.87)	13.006 (18.28)	13.006 (18.08)	13.007 (11.25)	13.007 (12.26)	13.007 (12.27)	13.006 (18.08)	13.006 (18.08)	13.006 (18.08)
$(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$	–	8.006 (19.22)	8.011 (6.30)	–	–	–	–	–	–
$(X_1 - \bar{X}_1)^2$	–	–	–0.012 (–0.01)	–	–	–	–0.012 (–0.01)	8.005 (10.79)	5.327 (8.88)
$(X_2 - \bar{X}_2)^2$	–	–	0.002 (0.00)	–	–	–	4.010 (10.81)	0.002 (0.00)	2.67 (8.91)
$X_1^\psi X_2^\psi$	–	–	–	5.477 (4.00)	–	–	–	–	5.345 (6.30)
$X_1^\psi (X_2 - \bar{X}_2)$	–	–	–	–	8.021 (7.42)	–	8.011 (6.30)	–	–
$(X_1 - \bar{X}_1) X_2^\psi$	–	–	–	–	–	8.045 (7.45)	–	8.011 (6.30)	–
R^2	0.81	0.93	0.93	0.82	0.85	0.85	0.93	0.93	0.93

Notes: True model is $Y = 3 X_1 + 5 X_2 + 8 X_1 X_2 + \epsilon$ 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 200 and the number of simulations is 10000. The averages of the estimated t -statistics are shown in parentheses.

Table 6: **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	-5.002 (-3.91)	-4.999 (-4.18)	-5.007 (-4.17)	-5.002 (-3.90)	-5.001 (-4.96)	-5.008 (-4.92)	-5.008 (-4.92)	-5.008 (-4.92)
X_2	13.013 (14.41)	13.009 (15.41)	13.016 (15.35)	13.013 (14.38)	13.012 (18.22)	13.011 (18.10)	13.011 (18.10)	13.011 (18.10)
$(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$	-	2.681 (5.40)	-	-	-	-	-	-
$(X_1 - \bar{X}_1)^2$	-	-	-	-	-	-2.665 (-4.47)	-7.994 (-7.70)	0.003 (-0.05)
$(X_2 - \bar{X}_2)^2$	-	-	-	-	-	2.660 (8.89)	3.991 (10.79)	0.001 (0.00)
$X_1^\psi X_2^\psi$	-	-	5.356 (5.37)	-	-	5.334 (6.29)	-	-
$X_1^\psi (X_2 - \bar{X}_2)$	-	-	-	-0.006 (-0.01)	-	-	7.991 (6.29)	-
$(X_1 - \bar{X}_1) X_2^\psi$	-	-	-	-	7.991 (11.01)	-	-	7.991 (6.29)
R^2	0.59	0.65	0.65	0.60	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); $X_1^\psi = M_2 = [I - P_{[constant, X_2]}] X_1$, $X_2^\psi = [I - P_{[constant, X_1]}] X_2$. A constant is included but not reported. The sample size is 200 and the number of simulations is 10000. The average of the estimated t -statistics are shown in parentheses.

Table 7: **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	0.999 (8.06)	2.994 (23.22)	2.997 (17.40)	2.996 (21.20)	2.997 (17.46)
X_2	–	0.002 (0.03)	0.002 (0.02)	0.001 (0.01)	0.002 (0.02)
X_1^2	1.000 (19.61)	–	–	–	–
$(X_1 - \bar{X}_1)(X_2 - \bar{X}_2)$	–	0.665 (12.44)	–	–	–
$X_1^\psi X_2^\psi$	–	–	0.011 (0.06)	–	–
$X_1^\psi (X_2 - \bar{X}_2)$	–	–	–	1.000 (9.80)	–
$(X_1 - \bar{X}_1) X_2^\psi$	–	–	–	–	0.000 (–0.00)
R^2	0.92	0.86	0.76	0.84	0.76

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 200 and the number of simulations is 10000. The average of the estimated t -statistics are shown in parentheses.

Table 8: **Replication of Rajan and Zingales (1998): Table 4 (column 5)**

Dependent Variable: Annual compounded growth rate

	(1) [†]	(2)
I	-4.33 (-3.20)	-4.33 (-3.20)
ET	0.12 (0.82)	-
EA	1.33 (3.74)	-
$(E - \bar{E})T^\psi$	-	0.38 (2.40)
$(E - \bar{E})A^\psi$	-	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 coefficient estimates are not reported. All the coefficients are multiplied by 10. t -statistics are in parenthesis. 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 9: **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 (1.76)	-212.00 (-1.37)	362.44 (3.49)	318.03 (0.72)	354.22 (3.34)	370.00 (3.46)	347.27 (3.37)
O	35.36 (3.86)	-0.91 (-0.08)	44.17 (6.26)	69.00 (2.30)	41.05 (5.86)	40.65 (5.85)	44.59 (6.27)
I^2	-	-	-	0.07 (0.02)	-	-	-
O^2	-	-	-	-10.58 (-0.87)	-	-	-
IO	-	919.95 (3.79)	-	-	-	-	-
$(I - \bar{I})(O - \bar{O})$	-	-	919.95 (3.79)	743.35 (2.27)	-	-	-
$I^\psi O^\psi$	-	-	-	-	957.24 (3.61)	-	-
$I^\psi(O - \bar{O})$	-	-	-	-	-	950.65 (3.64)	-
$(I - \bar{I})O^\psi$	-	-	-	-	-	-	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 in 1913 divided by GDP in 1913. Coefficient estimates for per capita industrialization and its interaction with openness are multiplied by 1000. A constant is included to all of the regressions but not reported. t -statistics are in parenthesis. 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 10: 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)	(5)
<i>LRGDPW67</i>	-9.40 (-2.30)	-8.79 (-2.02)	155.84 (1.91)	13.78 (-2.80)	7.00 (-1.70)
<i>One share – one vote (OV)</i>	1.57 (0.13)	7.71 (1.52)	2.28 (0.41)	9.60 (2.01)	8.14 (1.82)
<i>Antidirector rights (AR)</i>	1.99 (0.78)	-12.79 (-1.20)	-9.78 (-0.96)	-2.28 (-1.32)	-1.98 (-1.22)
<i>Creditor Rights (CR)</i>	9.98 (2.64)	5.19 (0.77)	6.19 (0.96)	1.79 (0.92)	2.05 (1.15)
<i>Rule of Law (RL)</i>	3.67 (1.42)	6.51 (1.25)	3.41 (0.66)	4.20 (2.89)	2.07 (1.71)
<i>RCT</i>	47.71 (1.49)	-36.03 (-0.99)	-30.62 (-0.88)	0.38 (0.06)	2.92 (0.46)
<i>LRGDPW67</i> ²	-	-	-9.25 (-2.02)	-	-
<i>AR</i> ²	-	1.81 (1.07)	1.47 (0.91)	-	-
<i>CR</i> ²	-	-0.77 (-0.49)	-0.87 (-0.58)	-	-
<i>RL</i> ²	-	-0.30 (-0.77)	0.02 (0.04)	-	-
<i>RCT</i> ²	-	36.49 (1.13)	27.76 (0.90)	-	-
<i>OV RCT</i>	9.10 (0.60)	-	-	-	-
<i>AR RCT</i>	-5.30 (-1.42)	-	-	-	-
<i>CR RCT</i>	-10.29 (-2.17)	-	-	-	-
<i>RL RCT</i>	-1.01 (-0.34)	-	-	-	-
$(OV - \overline{OV})(RCT - \overline{RCT})$	-	-9.94 (-0.57)	3.36 (0.19)	-	-
$(AR - \overline{AR})(RCT - \overline{RCT})$	-	-7.29 (-1.64)	-5.78 (-1.35)	-	-
$(CR - \overline{CR})(RCT - \overline{RCT})$	-	-11.14 (-2.30)	-3.73 (-0.64)	-	-
$(RL - \overline{RL})(RCT - \overline{RCT})$	-	1.22 (0.35)	2.51 (0.73)	-	-
$OV^\psi (RCT - \overline{RCT})$	-	-	-	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)
<i>R</i> ²	0.50	0.59	0.64	0.50	0.53

Notes: The column marked with † replicates Castro, Clementi, and MacDonald (2004). *RCT* is a variable created to measure restrictions on capital transactions. 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. *LRGDPW67* is the natural logarithm of real gross domestic product per worker in 1967. Finally, the index rule of law, *RL*, proxies for the quality of law enforcement. The coefficients are multiplied by 1000. Sample size is 43. A constant is included to all of the regressions but not reported. *t*-statistics are in parenthesis. 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 11: Replication of Caprio, Laeven, and Levine (2007): Table 5 (Column 1)

Dependent Variable: Market-to-book

	(1) [†]	(2)	(3)	(4)
<i>Loan Growth</i>	0.27 (0.76)	0.64 (1.42)	0.19 (0.54)	0.19 (0.53)
<i>Rights</i>	0.31 (5.75)	-0.22 (-1.10)	0.14 (3.43)	0.07 (1.81)
<i>CF</i>	2.27 (4.33)	-1.57 (-2.98)	-0.57 (-3.23)	-0.44 (-2.35)
<i>Loan Growth</i> ²	0.27 (0.76)	-1.17 (-1.44)	-	-
<i>Rights</i> ²	-	0.05 (1.51)	-	-
<i>CF</i> ²	-	1.34 (2.03)	-	-
<i>CF Rights</i>	-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)
<i>R</i> ²	0.20	0.23	0.22	0.15

Notes: The column marked with † replicates Caprio, Laeven, and Levine (2007). A constant is included to all of the 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 are in parenthesis. 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 12: Replication of Easterly, Levine, and Roodman (2004): Table 1 (Column 1)

Dependent Variable: Growth of GDP per capita

	(1) [†]	(2)	(3)	(4)
<i>Aid</i>	-0.02 (-0.12)	-0.01 (-0.02)	0.24 (1.49)	0.18 (1.20)
<i>Policy (P)</i>	0.71 (2.92)	0.77 (3.30)	0.94 (4.82)	0.98 (5.30)
<i>Ethnic (Eth)</i>	-0.42 (-0.53)	-2.25 (-0.75)	0.02 (0.02)	-0.07 (-0.09)
<i>Assassinations (Ass)</i>	-0.45 (-1.49)	-0.33 (-0.87)	-0.10 (-0.66)	-0.09 (-0.59)
<i>Aid</i> ²	-	0.04 (0.85)	-	-
<i>P</i> ²	-	0.13 (1.65)	-	-
<i>LGDP</i> ²	-	-0.56 (-1.71)	-	-
<i>Eth</i> ²	-	2.76 (0.75)	-	-
<i>Ass</i> ²	-	0.03 (0.74)	-	-
<i>IQ</i> ²	-	-0.21 (-1.39)	-	-
<i>M2/GDP lagged</i> ²	-	-0.00 (-0.34)	-	-
<i>Aid P</i>	0.19 (1.84)	-	-	-
<i>Eth Ass</i>	0.79 (1.28)	-	-	-
$(Aid - \overline{Aid})(P - \overline{P})$	-	0.23 (2.07)	-	-
$(Eth - \overline{Eth})(Ass - \overline{Ass})$	-	0.64 (0.94)	-	-
$Aid^\psi (P - \overline{P})$	-	-	0.19 (1.18)	-
$Eth^\psi (Ass - \overline{Ass})$	-	-	1.40 (1.69)	-
$(Aid - \overline{Aid}) P^\psi$	-	-	-	0.23 (2.34)
$(Eth - \overline{Eth}) Ass^\psi$	-	-	-	0.80 (1.22)
<i>R</i> ²	0.39	0.41	0.39	0.40

Notes: The column marked with † replicates Easterly, Levine, and Roodman (2004). Data is for 1970-1993 and the sample size is 270. *Aid* is development assistance divided by real GDP. *Policy (P)* is a regression-weighted average of macroeconomic policies. *LGDP* is logarithm of initial GDP per capita. *Ethnic (Et)* is ethnic fractionalization. *Assassinations (Ass)* is per million population. *Sub-Saharan Africa (SSA)* and *Fast – growing E.Asia (FEA)* are dummy variables. *Institutional quality (IQ)* is an index of institutional quality. *M2/GDP lagged* is a measure of financial depth. *t*-statistics are in parenthesis. All the regressors are included but not displayed for: a constant, *LGDP*, *SSA*, *FEA*, *IQ*, and *M2/GDP lagged*. The new variables which are created according to the Frisch-Waugh theorem:

$$Aid^\psi = (I - P_{[c, LGDP, P, GDP, Et, Ass, SSA, FEA, IQ, M2/GDP\ lagged, perioddummies]}) Aid,$$

$$P^\psi = (I - P_{[c, LGDP, Aid, GDP, Et, Ass, SSA, FEA, IQ, M2/GDP\ lagged, perioddummies]}) P,$$

$$Et^\psi = (I - P_{[c, LGDP, Aid, P, GDP, Ass, SSA, FEA, IQ, M2/GDP\ lagged, perioddummies]}) Et,$$

$$Ass^\psi = (I - P_{[c, LGDP, Aid, P, GDP, Et, SSA, FEA, IQ, M2/GDP\ lagged, perioddummies]}) Ass.$$

Table 13: **Replication of Spilimbergo (2009): Table 2a (Column 2)**

Dependent Variable: Democracy Measure: *Polity2* Levels

	(1) [†]	(2)	(3)
<i>Democracy</i> _{<i>t-5</i>}	0.45 (9.61)	0.44 (8.46)	0.44 (8.44)
<i>Students Abroad</i> _{<i>t-5</i>} (<i>S</i>)	24.23 (2.81)	24.23 (2.55)	-1.82 (-0.39)
<i>Democracy in host countries</i> _{<i>t-5</i>} (<i>DH</i>)	0.12 (2.23)	0.12 (2.23)	0.10 (1.84)
<i>S</i> _{<i>t-5</i>} <i>DH</i> _{<i>t-5</i>}	-33.71 (-2.71)	-33.31 (-2.47)	-
$(S_{t-5} - \overline{S_{t-5,i}})(DH_{t-5} - \overline{DH_{t-5,i}})$	-	-	56.44 (1.73)
<i>Time effects</i>	Yes	Yes	Yes
<i>Country effects</i>	Yes	Yes	Yes
<i>Observations</i>	1107	1121	1121
<i>R</i> ²	0.41	0.82	0.82

Notes: The column marked with † replicates Spilimbergo (2009). The data forms an unbalanced panel comprised at five year intervals between 1955 and 2000. The dependent variable, *Polity2*, is the composite Polity II democracy index from the Polity IV data set which is the difference between Polity's democracy and autocracy indices. *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.