Diplomats, Quo Vadis?

The determinants of Brazilian diplomatic presence

ANNEX

Contents

Legislation consulted 2

Empirical operationalization of determinants 3

External and structural factors 3

Special relations status 4

Number of agreements 4

Number of Brazilian residents 4

Group membership 5

Bureaucratic factors 5

Post tradition 5

Language skill 7

R Script 8

Additional materials 14

Legislation consulted

Decreto nº 93.325 from 01/10/1986. Available at:

<http://www.planalto.gov.br/ccivil_03/decreto/Antigos/D93325.htm>. Accessed on

18-05-2019.

Decreto n.º 5.979 from 06-12-2006. Available at: <

http://www.planalto.gov.br/ccivil_03/_Ato2015-

2018/2016/Decreto/D8817.htm#art10>. Accessed on 20-05-2017.

Decreto nº 9.683 from 09-01-2019. Available at:

<http://www.planalto.gov.br/ccivil_03/_Ato2019-2022/2019/Decreto/D9683.htm>.

Accessed on 01-05-2019.

Lei nº 11.440, from 29-12-2006. Available at: <

http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/L11440.htm> Accessed

on 06-06-2017.

MRE's internal regiment from 2008 (*Portaria nº 212, 30-04-2008*). Available at:

<http://www.itamaraty.gov.br/images/RISE.pdf> Accessed on 14-04-2019.

Empirical operationalization of determinants

External and structural factors

Values for partner-country GDP were obtained from the World Bank. Values are in constant 2010 USD.

Total bilateral trade was calculated by summing imports and exports between Brazil and the partner country. Data were obtained from the UNComtrade database.[1] Values are in constant 2010 USD.

Data on how many IGOs are headquartered in the country were obtained from the Yearbook of International Organizations. From the Yearbook's total count of IGOs, we retained only those referring to intergovernmental entities (i.e.: no NGOs), for which confirmed information was available. [2]

Geographical data on contiguity and distance between capitals was obtained from the CEPII dataset (Head et al. 2010, as referenced in main article).

[1] Available at: https://comtrade.un.org/data. Accessed on 09-02-2019

[2] That is, based on the Yearbook's categorization, only Type II = "g" entries were retained, and Type I = "H", "J", or "U" were discarded. Available at: https://uia.org/ybio/. Accessed on 28-09-2019

Special relations status

Number of agreements

It is particularly challenging to operationalize the concept of strategic partnerships, due in part to its overuse. The annals of Brazilian international agreements testify to the malleability of this label: Brasília has "partnerships" with some countries[3], "alliances" with others[4], "dialogues"[5] or even "mechanisms"[6] with others more – all of them "strategic" in name.

In light of such variation, our operationalization strategy cannot rely on verbatim mentions of the term "strategic partnership" alone. As special relations are expected to result of prolonged and dense interactions, a consistent record of bilateral agreements should be their most valid indicator. Therefore, in order to establish if any country had special relations with Brazil, we counted the number of bilateral agreements signed between them, as recorded at MRE's online repository, *Concordia.*[7] For this study, we have considered all agreements starting in 1900, except those classified under the "replaced" status.[8] In total, 5,567 agreements were considered.

Number of Brazilian residents

In order to count the number of Brazilians living abroad, we consulted a series of population estimates published by the MRE between 2008 and 2015.[9] Though these

[3] E.g.: Russia (2002), European Union (2007).
[4] E.g.: Peru (2003), Venezuela (2005)
[5] E.g.: US (2005), Canada (2012)
[6] E.g.: Nigeria (2013)
[7] Available at: https://concordia.itamaraty.gov.br/
[8] According to the MRE, the "replaced" (*substituído*) status implies that another agreement was edited in place of the previous one. These entries should, then, be ignored in order to avoid double counting.
[9] Available at: http://www.brasileirosnomundo.itamaraty.gov.br/a-comunidade/estimativas-populacionais-das-comunidades. Accessed on 16-02-19

figures are highly approximative, they provide the best available information on the Brazilian expatriate community at present.

Group membership

The following countries were coded as pertaining to the groups considered (G7, BRICS, Mercosul and CPLP).

Group	Members
G7	Canada, France, Germany, Italy, Japan, United Kingdom, United States
BRICS (from 2009 on)	Russia, India, China, South Africa (from 2011 on)
Mercosul	Argentina, Paraguay, Venezuela (from 2012 on), Uruguay
CPLP	Angola, Cape Verde, East Timor, Equatorial Guinea (from 2014 on), Guinea-Bissau, Mozambique, Portugal, Sao Tome and Principe

Bureaucratic factors

Post tradition

Post tradition was measured through the opening year of embassies abroad. It was considered that older posts are more traditional and, therefore, more coveted by MRE employees. We have gathered information on embassies and consulates, while legations and honorary stations were discarded. In total, we had information on approximately 640 city-level records. It was not possible to retrieve this information for 10 stations. Based on the date when an embassy was first opened or upgraded,

countries were ranked on a 12-point scale. Thresholds for the scale were based on decades, so that posts opened in the 2010s were ranked 1 and those prior to 1910 were ranked 12, as summarized below:

Opening prior to:	1910	1920	1930	1940	1950	1960	1970	1980	1990	2000	2010	2015
Points on scale	12	11	10	9	8	7	6	5	4	3	2	1

Data were collected, initially, (1) from official websites of the Brazilian representations abroad, (2) from the *Portal Consular*[10] run by MRE, and (3) from Decrees published at the Federal Executive's website.[11] For instances when these sources were insufficient, the diplomatic stations were contacted directly via email. This was the case for approximately 50 coutnries. Lastly, ministerial reports were also consulted[12] – for the years 1830 to 1960 –, as well as the books *1808 – 2008: Dois Séculos de História da Organização do Itamaraty* (volumes 1 and 2), being both sources available through the FUNAG. Whenever there was a conflict in the dates reported in any of those information sources, we adopted the following criteria for disambiguation. The information on official websites of the representations were prioritized, followed by the *Portal Consular,* Decrees and, lastly, the FUNAG records and emails to the stations.

[10] Available at: <http://www.portalconsular.itamaraty.gov.br/>. Accessed on 13-04-19
11 Available at: < https://legislacao.planalto.gov.br/legisla/legislacao.nsf/fraWeb?OpenFrameSet&Frame=frmWeb2&Src=/legisla/legislacao.nsf%2FFrmConsultaWeb1%3FOpenForm%26AutoFramed>. Accessed on 13-04-19
12 Available at: < http://www.funag.gov.br/chdd/index.php/relatorios-do-ministerio>. Accessed on 13-04-19

Whenever there was territorial disintegration of a country (i.e.: Yugoslavia, Austro-Hungarian Empire and Czechoslovakia), the existing posts were considered extinct in the same date of the state's breakup. New records were set up for the states that emerged afterwards. The same took place for cases of political independence, decolonization or transfer of sovereignty, as in the cases of India, Jamaica and Lebanon. Finally, states for which there was no information on the activity of the diplomatic posts were removed from the analysis. In total, there 23 such instances, among them Kosovo, New Caledonia, Pitcairn and Réunion.

Language skill

Language skill measures whether MRE employees should be properly skilled to master the host country's language, based on the institutional training received. The variable is a dummy registering 1 for countries wherein the official language is one of the following four: Portuguese, English, French or Spanish. The first language is considered unproblematic as it corresponds to the native language of MRE employees. The remaining three are mandatory language courses offered to diplomats at the IRBR, so that it is expected that they complete training duly equipped in these languages. Information on the official languages of countries were obtained from the CEPII dataset (Head et al. 2010). In total, 142 countries had one of those four languages as an official language; while 81 did not.

R Script

```
#### REPLICATION SCRIPT: QUO VADIS DIPLOMATS

# libraries
library(foreign)
library(tidyverse)
library(broom)
library(reshape2)
library(car)
library(Hmisc)
library(plm)
library(stargazer)
library(MASS)
library(vcd)
library(AER)

# Codebook
##  iso3: ISO-3 country code
##  rankmean: post hierarchial ranking (a-d), national mean ('rank' variable in
csRank is re-scaled 1-4 integer)
##  spRelct: count of number of bilateral agreements with BR
##  undp_hdim: mean HDI score
##  nPers: number of personnel
##  contig: dummy for geogr. contiguity
##  distw: distance between capitals
##  n_igos: number of intergovernmental organizations HQs
##  nBrRes: number of Brazilian residents
##  emb12: country score in post tradition (12 point scale)
##  lang_ob: dummy for official language PT, ENG, FR or SPA
##   code for transformations: _ln (log()), _bi (billions), _l1 _l2 (lagged
value)
##  t: time trend



# SECTION 4.2: DETERMINANTS OF POST RANKING

# Cross-section with mean 2008-2015 values for post ranking and MRE criteria
csRank <- read.csv(file="CSrank.csv")
csRank$rank <- as.factor(csRank$rank)

# Ordinal Logistic Regression
## formulas
rankForm <- rank ~ gdp_bi_ln + spRelct_ln + undp_hdim_ln
rankForms <- rank ~ scale(gdp_bi_ln) + scale(spRelct_ln) + scale(undp_hdim_ln)

## regressions
polr1 <- polr(rankForm, data=csRank, Hess=T) #un-standardized values
summary(polr1)
pval1 <- pnorm(abs(coef(summary(polr1))[, "t value"]), lower.tail = FALSE) * 2
cbind(coef(summary(polr1)), "p value" = round(pval1, 3))
confint(polr1)
### OR and CI
or1 <- exp(cbind(OR = coef(polr1), confint(polr1)))
or1

polr1s <- polr(rankForms, data=csRank, Hess=T) # standardized independent
variables (replicates Table 3)
summary(polr1s)
pval1s <- pnorm(abs(coef(summary(polr1s))[, "t value"]), lower.tail = FALSE) *
2
cbind(coef(summary(polr1s)), "p value" = round(pval1s, 3))
confint(polr1s)
```

```
### OR and CI
or1s <- exp(cbind(OR = coef(polr1s), confint(polr1s)))
or1s

# Plots (replicates Figure 3)
## Histograms for empirical count of cases
ggplot(csRank, aes(x=undp_hdim_ln, fill=rank))+
  geom_histogram(position = "identity", color="black")+
  scale_fill_manual(values=c("black", "gray40", "gray80", "gray99"))+
  ylim(0,10)+
  labs(title="HDI histogram", x="HDI (log)", y="Count", fill="Observed\nRank")

ggplot(csRank, aes(x=gdp_bi_ln, fill=rank))+
  geom_histogram(position = "identity", color="black")+
  scale_fill_manual(values=c("black", "gray40", "gray80", "gray99"))+
  ylim(0,10)+
  labs(title="GDP     histogram",     x="GDP     (bi,     log)",     y="Count",
fill="Observed\nRank")
dev.off()

ggplot(csRank, aes(x=spRelct_ln, fill=rank))+
  geom_histogram(position = "identity", color="black")+
  scale_fill_manual(values=c("black", "gray40", "gray80", "gray99"))+
  ylim(0,10)+
  labs(title="N.   Agreem.   histogram",   x="N.   Agreem.   (log)",   y="Count",
fill="Observed\nRank")

## Probability curves
plotdata1 <- csRank[complete.cases(csRank[,c(5,7,9)]),c(5,7,9)]
plotdata1 <- scale(plotdata1)
plotdata1 <- as.data.frame(plotdata1)
xbgdp <- plotdata1$gdp_bi_ln*coef(polr1s)[1]
xbsp <- plotdata1$spRelct_ln*coef(polr1s)[2]
xbhdi <- plotdata1$undp_hdim_ln*coef(polr1s)[3]

logistic_cdf <- function(x) {
  return( 1/(1+exp(-x) ) )
}

plotdatag <- as.data.frame(plotdata1$gdp_bi_ln)
colnames(plotdatag)[1] <- "gdp_bi_ln"
plotdatag$D <- logistic_cdf( -0.7578 - xbgdp )
plotdatag$C <- logistic_cdf( 3.5624 - xbgdp ) - logistic_cdf( -0.7578 - xbgdp )
plotdatag$B <- logistic_cdf( 6.4487 - xbgdp ) - logistic_cdf( 3.5624 - xbgdp )
plotdatag$A <- 1 - logistic_cdf( 3.5624 - xbgdp )

ggplot(melt(plotdatag, id.vars = c(colnames(plotdatag)[1]),
            variable.name = "predRank", value.name = "prob"),
       aes(x=gdp_bi_ln,                                                    y=prob,
linetype=predRank))+geom_line(aes(linetype=predRank), size=1.2)+
  scale_linetype_manual(values=c("solid", "longdash", "dotdash", "dotted"))+
 ylim(c(0,1))+
  labs(title="GDP (log) x Probability", x="GDP (bi, log)", y="Probability",
linetype="Predicted\nRank")

plotdatas <- as.data.frame(plotdata1$spRelct_ln)
colnames(plotdatas)[1] <- "spRelct_ln"
plotdatas$D <- logistic_cdf( -0.7578 - xbsp )
plotdatas$C <- logistic_cdf( 3.5624 - xbsp ) - logistic_cdf( -0.7578 - xbsp )
plotdatas$B <- logistic_cdf( 6.4487 - xbsp ) - logistic_cdf( 3.5624 - xbsp )
plotdatas$A <- 1 - logistic_cdf( 3.5624 - xbsp )

ggplot(melt(plotdatas, id.vars = c(colnames(plotdatas)[1]),
            variable.name = "predRank", value.name = "prob"),
       aes(x=spRelct_ln,                                                   y=prob,
linetype=predRank))+geom_line(aes(linetype=predRank), size=1.2)+
  scale_linetype_manual(values=c("solid", "longdash", "dotdash", "dotted"))+
```

```
  ylim(c(0,1))+
  labs(title="N. Agreements (log) x Probability", x="N. Agreem. (log)",
y="Probability", linetype="Predicted\nRank")

plotdatah <- as.data.frame(plotdata1$undp_hdim_ln)
colnames(plotdatah)[1] <- "undp_hdim_ln"
plotdatah$D <- logistic_cdf( -0.7578 - xbhdi )
plotdatah$C <- logistic_cdf( 3.5624 - xbhdi ) - logistic_cdf( -0.7578 - xbhdi )
plotdatah$B <- logistic_cdf( 6.4487 - xbhdi ) - logistic_cdf( 3.5624 - xbhdi )
plotdatah$A <- 1 - logistic_cdf( 3.5624 - xbhdi )

ggplot(melt(plotdatah, id.vars = c(colnames(plotdatah)[1]),
            variable.name = "predRank", value.name = "prob"),
       aes(x=undp_hdim_ln,                                                y=prob,
linetype=predRank))+geom_line(aes(linetype=predRank), size=1.2)+
  scale_linetype_manual(values=c("solid", "longdash", "dotdash", "dotted"))+
  ylim(c(0,1))+
  labs(title="HDI (log) x Probability", x="HDI (log)", y="Probability",
linetype="Predicted\nRank")


# SECTION 4.3: DETERMINANTS OF DIPLOMATIC PRESENCE

# Cross-section with mean 2008-2015 values
cs <- read.csv(file="CSmean.csv")

# OLS regression
## formulas
olsForm1 <- nPers_ln ~ contig + distw + gdp_bi_ln + trade_ln + n_igos +
  spRelct_ln + nBrRes_ln + g7 + brics + mercosul + cplp +
  emb12 + rankmean + lang_ob # full model
olsForm1s <- scale(nPers_ln) ~ scale(contig) + scale(distw) + scale(gdp_bi_ln)
+ scale(trade_ln) + scale(n_igos) +
  scale(spRelct_ln) + scale(nBrRes_ln) + scale(g7) + scale(brics) +
scale(mercosul) + scale(cplp) +
  scale(emb12) + scale(rankmean) + scale(lang_ob) # same as 1, both Y and X
standardized
olsForm1ss <- nPers_ln ~ scale(contig) + scale(distw) + scale(gdp_bi_ln) +
scale(trade_ln) + scale(n_igos) +
  scale(spRelct_ln) + scale(nBrRes_ln) + scale(g7) + scale(brics) +
scale(mercosul) + scale(cplp) +
  scale(emb12) + scale(rankmean) + scale(lang_ob) # same as 1, only X
standardized

olsForm2 <- nPers_ln ~ contig + distw + trade_ln + n_igos +
  spRelct_ln + nBrRes_ln + g7 + brics + mercosul + cplp +
  emb12 + lang_ob # reduced model, without gdp and rank

## regressions
olsMean <- lm(data=cs, olsForm1)
summary(olsMean) # regular OLS regression (Model 1)

olsMeans <- lm(data=cs, olsForm1s)
summary(olsMeans) # standardized coefficients, X and Y scaled

olsMeanss <- lm(data=cs, olsForm1ss)
summary(olsMeanss) # standardized coefficients, only X scaled

olsMean2 <- lm(data=cs, olsForm2) # reduced OLS regression (Model 2)

## diagnostics
qqPlot(olsMean)
hist(olsMean$residuals)
residualPlot(olsMean)
vif(olsMean) # gdb_bi_ln exceeds vif=10
infIndexPlot(olsMean, vars=c("Cook", "Studentized", "hat"), id.n = 3)
outlierTest(olsMean) # US does not qualify as outlier
```

```
bptest(olsMean) # no heteroskedasticity
ncvTest(olsMean) # error variance does not change with fitted-value range


# Negative binomial regression
csi <- cs
csi$nPers <- as.integer(csi$nPers) # transform into integer for nb regression

## Test for overdispersion
mean(csi$nPers)
var(csi$nPers) # 12.4 < 475.5. Overdispersed

## Poisson regression to test for overdispersion
poisMean <- glm(nPers ~ contig + distw + gdp_bi_ln + trade_ln + n_igos +
spRelct_ln +
                  nBrRes_ln + g7 + brics + mercosul + cplp + emb12 + rankmean +
                  lang_ob, family=poisson, data=csi) # same form as olsForm1
summary(poisMean)
### dispersion test
dispersiontest(poisMean, trafo = NULL, alternative="greater")

## Negative binomial regression
nbMean <- glm.nb(nPers ~ contig + distw +
                   gdp_bi_ln + trade_ln + n_igos +
                   spRelct_ln + nBrRes_ln +
                   g7 + brics + mercosul + cplp +
                   emb12 + rankmean + lang_ob,
                 data=csi)
summary(nbMean)

nbMeans <- glm.nb(nPers ~ scale(contig) + scale(distw) +
                    scale(gdp_bi_ln) + scale(trade_ln) + scale(n_igos) +
                    scale(spRelct_ln) + scale(nBrRes_ln) +
                    scale(g7) + scale(brics) + scale(mercosul) + scale(cplp) +
                    scale(emb12) + scale(rankmean) + scale(lang_ob),
                  data=csi) #  standardized independent variables
summary(nbMeans)

nbMean2 <- glm.nb(nPers ~ contig + distw +
                    trade_ln + n_igos + spRelct_ln + nBrRes_ln +
                    g7 + brics + mercosul + cplp +
                    emb12 + lang_ob,
                  data=csi) # reduced model (no gdp and no rankmean)
summary(nbMean2)



# Panel data 2008-2015
pd <- read.csv(file="PD.csv")
pd <- pdata.frame(pd, index=c("iso3", "year")) # create pdataframe object

## centered version (only independent variables are centered)
pdc <- scale(as.matrix(pd[complete.cases(pd[,-c(4,5,7,8)]),-c(1:3,6)]))
pdc <- as.data.frame(pdc)
pdc <- cbind(pdc, pd[complete.cases(pd[,-c(4,5,7,8)]),c(1:3,6)])
pdc <- pdata.frame(pdc, index=c("iso3", "year"))

# Pooled OLS regression
## formulas
pdOlsForm1 <- nPers_ln ~ contig + distw +
  gdp_bi_ln + trade_ln + n_igos +
  spRelct_ln + nBrRes_ln +
  g7 + brics + mercosul + cplp +
  rankmean + lang_ob + emb12 + t + nPers_l1_ln # w/ lag1

pdOlsForm2 <- nPers_ln ~ contig + distw +
  gdp_bi_ln + trade_ln + n_igos +
```

```
  spRelct_ln + nBrRes_ln +
  g7 + brics + mercosul + cplp +
  rankmean + lang_ob + emb12 + t + nPers_l1_ln + as.factor(year) # w/ lag1 and
year effects

pdOlsForm3 <- nPers_ln ~ contig + distw +
  gdp_bi_ln + trade_ln + n_igos +
  spRelct_ln + nBrRes_ln +
  g7 + brics + mercosul + cplp +
  rankmean + lang_ob + emb12 + t + nPers_l1_ln + nPers_l2_ln + as.factor(year)
# w/ lag1+2 and year effects

pdOlsForm4 <- nPers_ln ~ contig + distw +
  trade_ln + n_igos +
  spRelct_ln + nBrRes_ln +
  g7 + brics + mercosul + cplp +
  lang_ob + emb12 + t + nPers_l1_ln + nPers_l2_ln + as.factor(year) # same as
3, reduced (no gdp and rank)



## Regressions
pdOls1 <- plm(pdOlsForm1, data=pd, model = "pooling")

## Test for unobserved heterogeneity
pwtest(pdOls1) # no individual effect
plmtest(pdOls1, effect="individual", type="bp") # no individual effect
plmtest(pdOls1, effect="time", type="bp") # significant time effects

## Regressions with time effects
pdOls2 <- plm(pdOlsForm2, data=pd, model = "pooling") # lag 1 + year effect
pdOls3 <- plm(pdOlsForm3, data=pd, model = "pooling") # lag1+2 + year effect
pdOls4 <- plm(pdOlsForm4, data=pd, model = "pooling") # reduced model

## Test for serial correlation
pwtest(pdOls2)
pbsytest(pdOls2, test="ar")
pbgtest(pdOls2)
pwtest(pdOls3)
pbsytest(pdOls3, test="ar")
pbgtest(pdOls3)  #  model  with  lag1+2  and  time  effects  presents  no  serial
correlation

## test for contemporaneous correlation
pcdtest(pdOls3, test="lm") # contemporaneous correlation is present

## bp test for homoskedasticity
bptest(pdOls3) # heteroskedasticity-robuts coefs required


# Export all 6 models side by side, original values (replicates Table 4)
coeftest
stargazer(olsMean, olsMean2, nbMean, nbMean2,
          coeftest(pdOls3, vcovHC(pdOls3, method="arellano")),
          coeftest(pdOls4, vcovHC(pdOls4, method="arellano")),
          decimal.mark = ".", no.space = T, align = T, digits = 2,
          type="html")   #  panel  data  models  use  heterosk-robust  coefs  via
coeftest



# PLOT STANDARDIZED COEFFICIENTS (Replicates Figure 4)
df.olsMeanss <- tidy(olsMeanss)
df.olsMeanss <- cbind(df.olsMeanss, confint_tidy(olsMeanss))
df.olsMeanss$model <- "OLS"
df.olsMeanss$term<- c("Intercept", "Contig.", "Distance", "GDP (log)", "Trade
(log)", "N. IGOs",
```

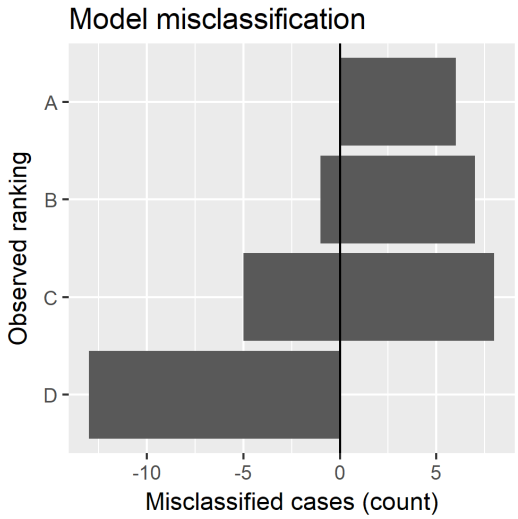
```
                      "N.Agreem.(log)",   "N.BR.Res.(log)",   "G7",   "BRICS",
"Mercosul","CPLP",
                      "Post Trad.", "Post Rank", "Lang.")

df.nbs <- tidy(nbMeans)
df.nbs <- cbind(df.nbs, confint_tidy(nbMeans))
df.nbs$model <- "Neg.Bin."
df.nbs$term  <-  c("Intercept",  "Contig.",  "Distance",  "GDP  (log)",  "Trade
(log)","N. IGOs",
                   "N.Agreem.(log)",     "N.BR.Res.(log)",     "G7",     "BRICS",
"Mercosul","CPLP",
                   "Post Trad.", "Post Rank", "Lang.")

pdOls5 <- plm(pdOlsForm3, data=pdc, model = "pooling") # same as pdOls3, but
using centered independent variables
df.pd <-tidy(coeftest(pdOls5, vcovHC(pdOls5, method="arellano")))
confint(coeftest(pdOls5, vcovHC(pdOls5, method="arellano")))
df.pd    <-    cbind(df.pd,    confint_tidy(coeftest(pdOls5,    vcovHC(pdOls5,
method="arellano"))))
df.pd$model <- "P.OLS"
df.pd$term <- c("Intercept", "Contig.", "Distance", "GDP (log)", "Trade (log)",
"N. IGOs",
                "N.Agreem.(log)",     "N.BR.Res.(log)",     "G7",     "BRICS",
"Mercosul","CPLP",
                "Post Rank", "Lang.", "Post Trad.", "t", "LDV1", "LDV2",
                "d2009", "d2010", "d2011", "d2012", "d2013", "d2014")



df.olsnbpd <- rbind(df.olsMeanss, df.nbs)
df.olsnbpd <- rbind(df.olsnbpd, df.pd)
df.olsnbpd.s <- df.olsnbpd[-c(49:54),]
df.olsnbpd.s$p <- ifelse(df.olsnbpd.s$p.value<0.05,1,0)
df.olsnbpd.s$order <- c(rep(1:15,2),1:12,14,15,13,16:18)
df.olsnbpd.s$term <- reorder(df.olsnbpd.s$term, -df.olsnbpd.s$order)

ggplot(df.olsnbpd.s, aes(x=term, y=estimate, shape=model, alpha=p))+
  geom_point(position=position_dodge(0.6), size=2)+
  geom_linerange(data=df.olsnbpd.s,          aes(ymin=conf.low,ymax=conf.high),
position=position_dodge(0.6))+
  geom_hline(yintercept = 0)+
  scale_shape_manual(name="Model", values=c("circle", "square", "triangle"))+
  scale_alpha(range = c(0.3,1), guide=F)+
  coord_flip()+
  theme(legend.position = c(0.8, 0.2))+
  xlab("Determinant")+ylab(expression(paste("Estimate (", italic("b"), ")")))


# INTERACTIVE MODEL (Annex)
pdOlsForm3i <- nPers_ln ~ contig + distw +
  gdp_bi_ln + trade_ln + n_igos +
  spRelct_ln + nBrRes_ln + nBrRes_ln*gdp_bi_ln +
  g7 + brics + mercosul + cplp +
  rankmean + lang_ob + emb12 + t + nPers_l1_ln + nPers_l2_ln + as.factor(year)
# same as 3, with interctive term

pdOls3i <- plm(pdOlsForm3i, data=pdc, model = "pooling") # on centered data

# Plot marginal effects
meplot <- function(model,var1,var2,int,vcov,ci=.95,
                   xlab=var2,ylab=paste("Marginal Effect of",var1),
                   main="Marginal Effect Plot",
                   me_lty=1,me_lwd=1,me_col="black",
                   ci_lty=1,ci_lwd=.5,ci_col="black",
                   yint_lty=2,yint_lwd=1,yint_col="black"){
  require(ggplot2)
  alpha <- 1-ci
```

```
  z <- qnorm(1-alpha/2)
  beta.hat <- coef(model)
  cov <- vcov
  z0                                                                    <-
seq(min(model.frame(model)[,var2],na.rm=T),max(model.frame(model)[,var2],na.rm
=T),length.out=1000)
  dy.dx <- beta.hat[var1] + beta.hat[int]*z0
  se.dy.dx    <-    sqrt(cov[var1,var1]    +    z0^2*cov[nrow(cov),ncol(cov)]    +
2*z0*cov[var1,ncol(cov)])
  upr <- dy.dx + z*se.dy.dx
  lwr <- dy.dx - z*se.dy.dx
  ggplot(data=NULL,aes(x=z0, y=dy.dx)) +
    labs(x=xlab,y=ylab,title=main) +
    geom_line(aes(z0, dy.dx),size = me_lwd,
              linetype = me_lty,
              color = me_col) +
    geom_line(aes(z0, lwr), size = ci_lwd,
              linetype = ci_lty,
              color = ci_col) +
    geom_line(aes(z0, upr), size = ci_lwd,
              linetype = ci_lty,
              color = ci_col) +
    geom_hline(yintercept=0,linetype=yint_lty,
               size=yint_lwd,
               color=yint_col)
}
meplot(model=pdOls3i,            var1="gdp_bi_ln",            var2="nBrRes_ln",
int="gdp_bi_ln:nBrRes_ln", vcov=vcov(pdOls3i),
       xlab="N. BR. Residents (log)", ylab="Marginal Effect of GDP bi (log)"
       )
```

Additional materials

The administration's convenience: misclassified stations in our logistic model

As mentioned in Section 4.2, our model misclassified a few of Brazil's posts abroad with regards to their ranking. The figure below aggregates these cases and reveals that overestimations (ranking predicted by the model was higher than true ranking) were more frequent for D posts. Negative values on the X axis count overestimations (predicted rank was higher than actual rank) and positive values underestimations (predicted ranks were lower than actual one).



Source: elaborated by the authors.

Extended version of Table 4 (including year dummies)

	Dependent variable:					
	nPers_ln *OLS*		nPers *negative binomial*		*coefficient test*	
	(1)	(2)	(3)	(4)	(5)	(6)
contig	0.91***	0.68***	0.66***	0.55***	0.19***	0.14***
	(0.18)	(0.16)	(0.18)	(0.15)	(0.05)	(0.05)
distw	-0.00	-0.00	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
gdp_bi_ln	0.08**		0.05		0.01	
	(0.04)		(0.05)		(0.01)	
trade_ln	0.01	0.02	0.06	0.06**	0.01	0.02***
	(0.02)	(0.02)	(0.04)	(0.03)	(0.01)	(0.01)
n_igos	0.001	0.003*	0.002*	0.002**	0.0002	0.0004
	(0.001)	(0.001)	(0.001)	(0.001)	(0.0005)	(0.0004)
spRelct_ln	0.01	0.004	0.07	0.07	0.01	0.01
	(0.04)	(0.04)	(0.05)	(0.05)	(0.01)	(0.01)
nBrRes_ln	0.09***	0.12***	0.13***	0.14***	0.02***	0.03***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
g7	0.71***	0.77***	0.47***	0.51***	0.16***	0.16***
	(0.16)	(0.17)	(0.15)	(0.15)	(0.05)	(0.05)
brics	0.74***	0.78***	0.62***	0.64***	0.17***	0.16***
	(0.17)	(0.18)	(0.17)	(0.17)	(0.04)	(0.04)

mercosul	0.77***	0.88***	0.55***	0.60***	0.11**	0.11**
	(0.22)	(0.23)	(0.16)	(0.16)	(0.05)	(0.05)
cplp	0.65***	0.43***	0.51***	0.40**	0.10**	0.07
	(0.16)	(0.15)	(0.17)	(0.16)	(0.05)	(0.04)
emb12	0.04**	0.06***	0.03	0.05***	0.01	0.01*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
rankmean	0.14**		0.06		0.02	
	(0.06)		(0.07)		(0.02)	
lang_ob	0.13*	0.02	0.20**	0.13	0.05*	0.03
	(0.08)	(0.07)	(0.09)	(0.09)	(0.03)	(0.03)
t					-0.01	-0.01
					(0.005)	(0.005)
nPers_l1_ln					0.60***	0.60***
					(0.04)	(0.04)
nPers_l2_ln					0.17***	0.18***
					(0.04)	(0.04)
as.factor(year)2009					-0.02	-0.02
					(0.03)	(0.03)
as.factor(year)2010					0.09***	0.08**
					(0.03)	(0.03)
as.factor(year)2011					0.06**	0.05*
					(0.03)	(0.03)
as.factor(year)2012					0.02	0.02
					(0.03)	(0.03)
as.factor(year)2013					0.03	0.03
					(0.03)	(0.03)
as.factor(year)2014					-0.01	-0.01
					(0.03)	(0.03)
Constant	0.56	0.36	-0.72	-0.66	-0.06	-0.14
	(0.42)	(0.30)	(0.64)	(0.42)	(0.15)	(0.11)
Observations	135	138	135	138		
R^2	0.90	0.88				
Adjusted R^2	0.88	0.87				
Log Likelihood			-317.70	-329.10		
theta			41.53*** (15.72)	36.90*** (13.63)		
Akaike Inf. Crit.			665.41	684.20		
Residual Std. Error	0.31 (df = 120)	0.32 (df = 125)				
F Statistic	73.97*** (df = 14; 120)	77.58*** (df = 12; 125)				
Note:						*p<0.1; **p<0.05; ***p<0.01

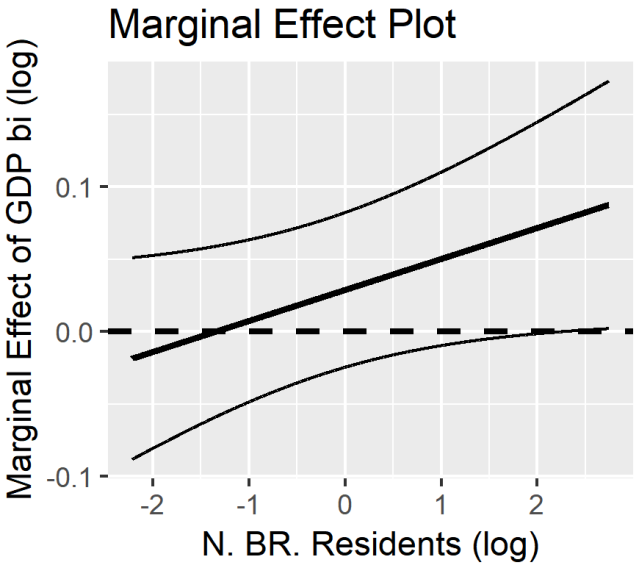
Source: elaborated by the authors.

Marginal Effects of N. Brazilian Residents x GDP

Some endogeneity is expected between the number of Brazilian residents and a country's GDP, as Brazilian expatriates will tend to prefer wealthy countries to live in. As an approximative attempt to disentangle the effect of a partner countries wealth from the demands of the local Brazilian community, we extended Model (5) to include an interaction term between both variables. The model was applied to a mean-centered version of the dataset so as to render interpretation more meaningful. The results are presented in the Table and Figure below and they suggest that the effect of the number of Brazilian residents on mission size is amplified as the partner GDP's grows.

	Dependent variable:
	nPers_ln
contig	0.056***
	(0.016)
distw	-0.011
	(0.016)
gdp_bi_ln	**0.029**
	(0.027)
trade_ln	0.029
	(0.020)
n_igos	-0.0002
	(0.013)
spRelct_ln	0.011
	(0.016)
nBrRes_ln	**0.076*** **
	(0.020)
g7	0.027*
	(0.014)
brics	0.027***
	(0.010)
mercosul	0.018
	(0.012)

cplp	0.026**
	(0.012)
rankmean	0.018
	(0.015)
lang_ob	0.019
	(0.012)
emb12	0.025
	(0.015)
t	-0.013
	(0.012)
nPers_l1_ln	0.545***
	(0.030)
nPers_l2_ln	0.155***
	(0.029)
as.factor(year)2009	-0.020
	(0.035)
as.factor(year)2010	0.086***
	(0.033)
as.factor(year)2011	0.057*
	(0.031)
as.factor(year)2012	0.023
	(0.030)
as.factor(year)2013	0.030
	(0.030)
as.factor(year)2014	-0.013
	(0.031)
gdp_bi_ln:nBrRes_ln	**0.021***
	(0.012)
Constant	2.066***
	(0.019)
Observations	925
R^2	0.920
Adjusted R^2	0.918
F Statistic	433.949*** (df = 24; 900)
Note:	*p<0.1; **p<0.05; ***p<0.01



Source: elaborated by the authors.