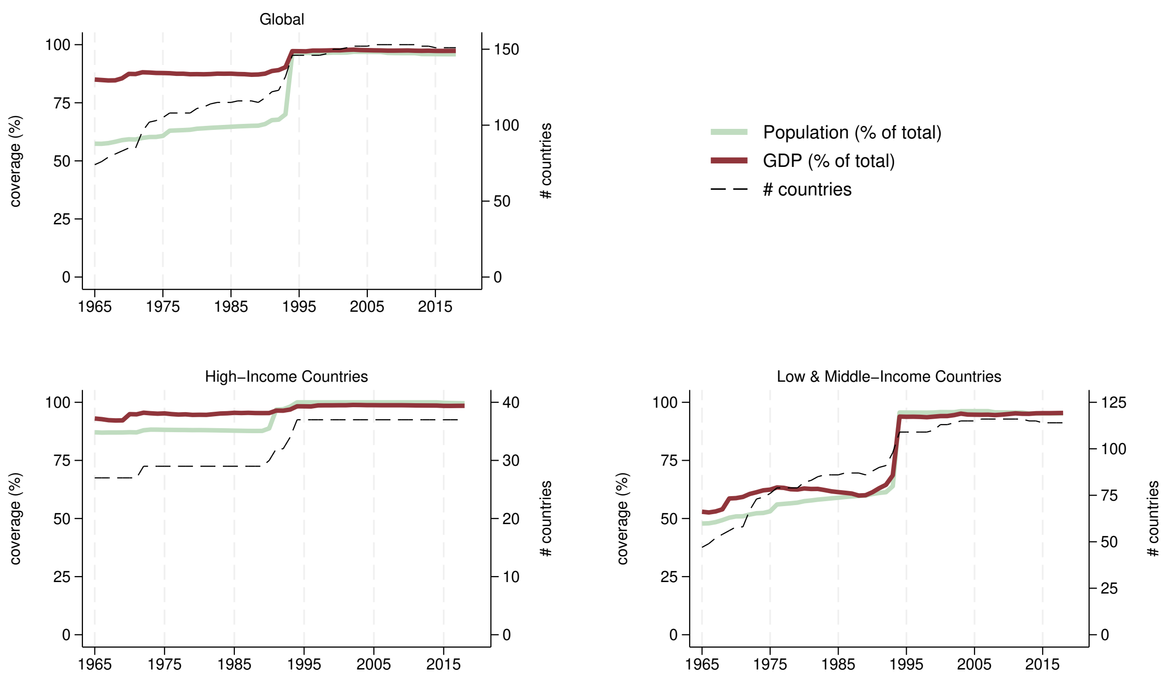
# Online Appendix

## *Globalization, Capital Taxation, and Development*

**Pierre Bachas Matthew Fisher-Post Anders Jensen Gabriel Zucman**

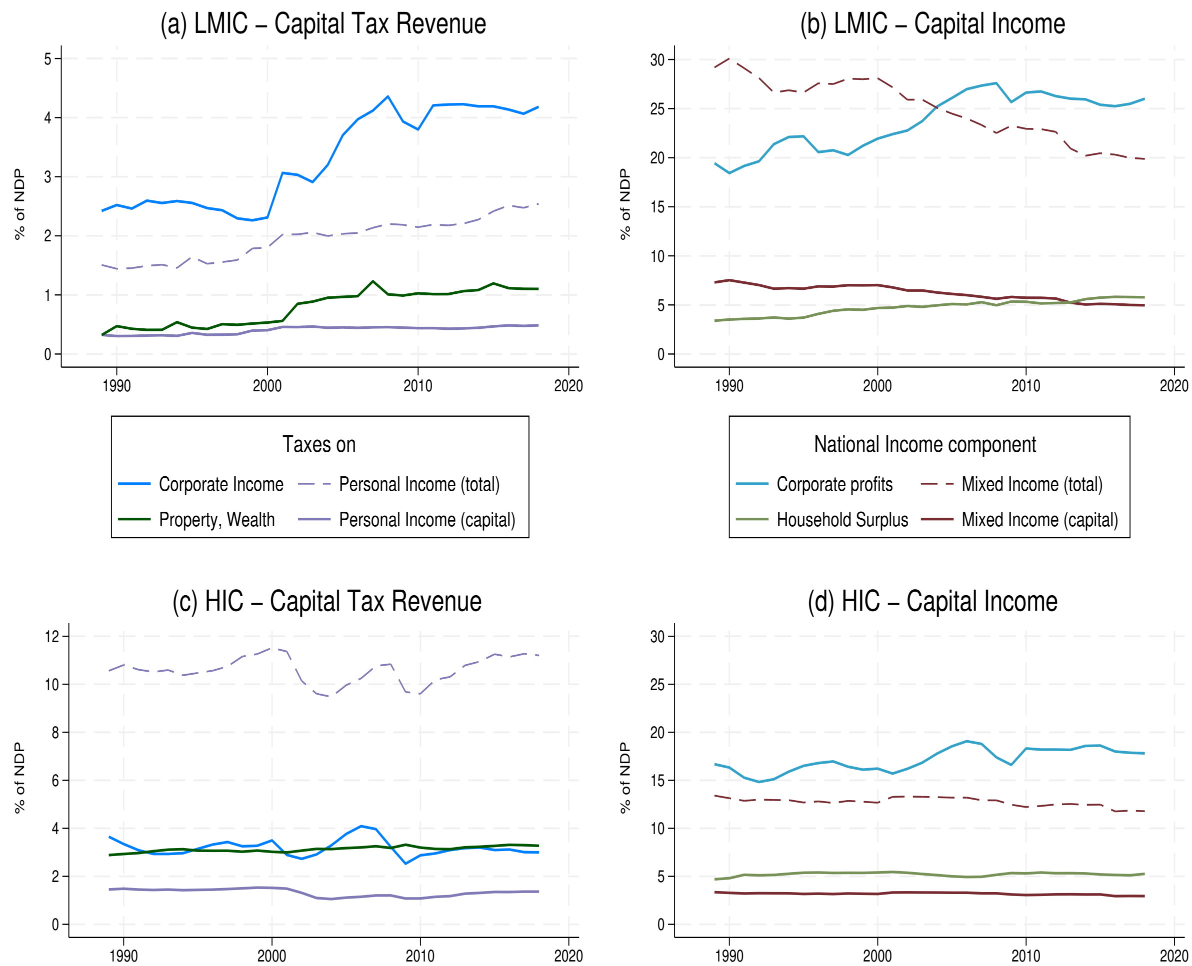
# Appendix A Additional Figures and Tables

Figure A1: Data Coverage of Effective Tax Rates



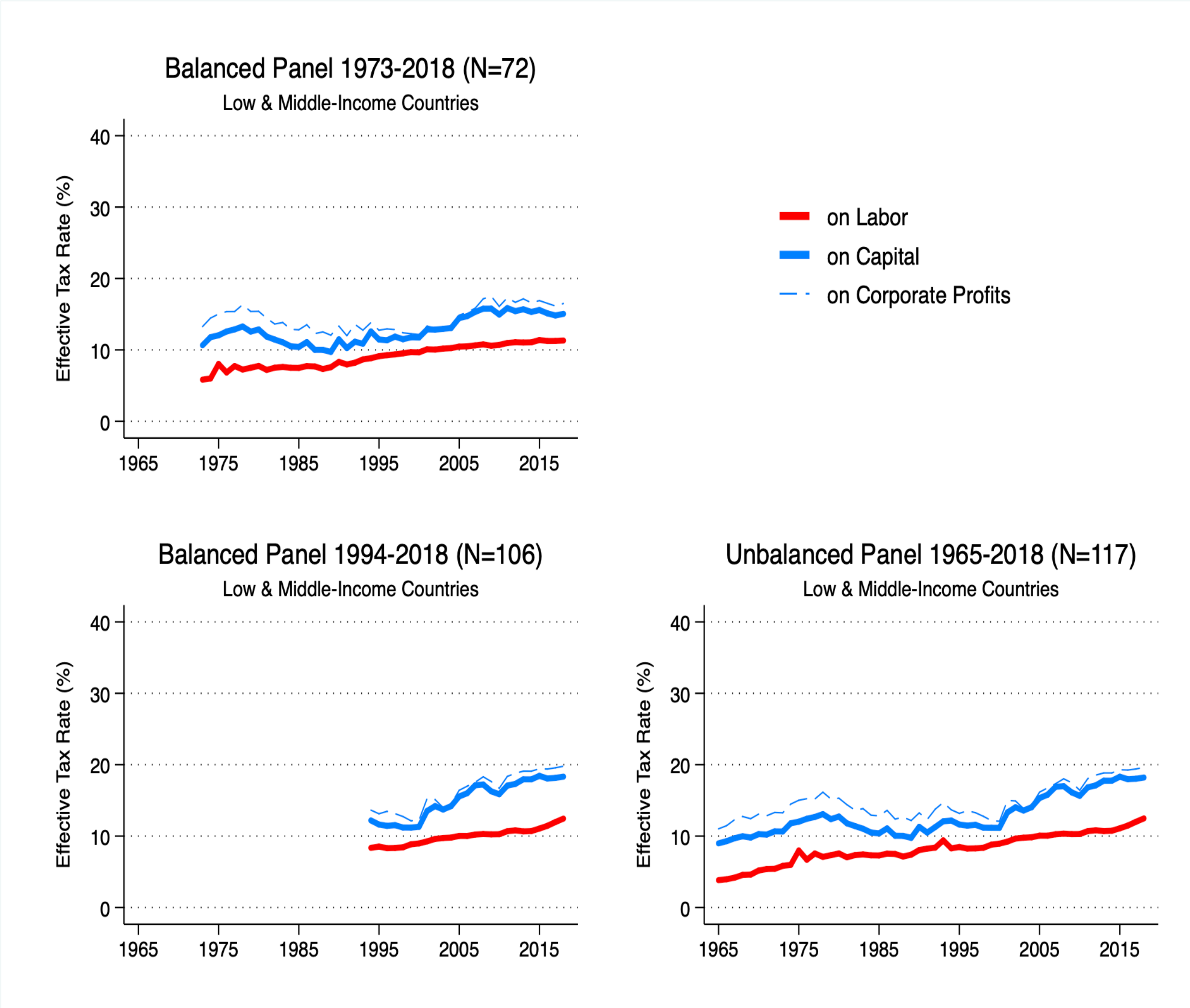

*Notes:* These panels show the coverage of our effective tax rate data between 1965 and 2018 at the global level (top left panel), in high income countries (bottom left panel), and in low- and middle-income countries (bottom right panel). The classification of countries into two development groups is defined in Section 3.3. The solid lines plot the percent of total population and GDP that are covered in our data (left axis). The dashed lines show the number of countries in the data (right axis). The dataset is composed of two quasi-balanced panels. The first covers the years 1965-1993 and excludes communist regimes. The second covers 1994-2018 and integrates former communist countries, in particular China and Russia. See Section 3.3 for more details.

Figure A2: Evolution of $ETR_K$ Components since 1989




*Notes:* These panels show the evolution of the components of $ETR_K$ between 1989 and 2018. This period is selected to match the period of rising $ETR_K$ in low and middle-income countries (LMICs). The left-hand side panels correspond to the taxes on capital (numerator of $ETR_K$): corporate income taxes; taxes on property, wealth and inheritance; and the share of personal income taxes allocated to capital (including capital gains and dividends). The right-hand side panels correspond to the national income components attributed to capital (denominator of $ETR_K$): corporate profits; operating surplus of households (rents); and the share of mixed-income attributed to capital. The top panels show LMICs, while the bottom panels show high-income countries (HICs). Series are weighted by countries' national domestic product in 2010. The tax revenue data between 1989-1993 for former command economies (e.g. China, Russia) is missing, and is imputed by assigning the 1994 values five years backward (between 1989-1993). The classification of countries into two development groups is defined in Section 3.3. This figure is discussed in Section 4.1.

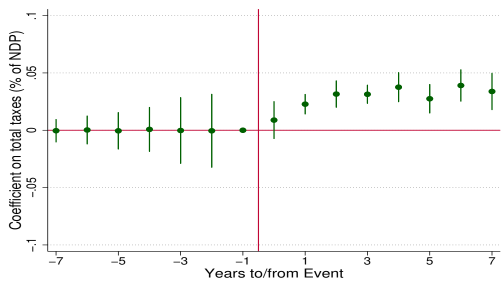
Figure A3: Effective Taxation in Developing Countries with Balanced Panel



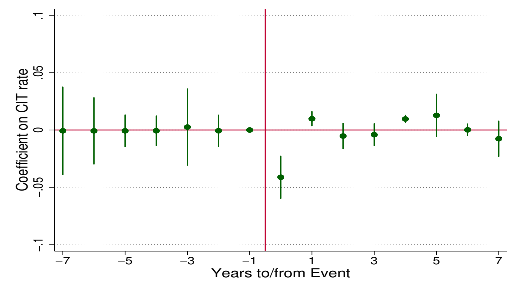

*Notes*: These panels show the evolution of effective taxation for three different samples of developing countries. The top left panel considers a balanced sample starting in 1973 until 2018 (N=72): this sample excludes ex-communist countries and notably China (whose series start in 1994). The bottom left panel considers a balanced sample of developing countries starting in 1994 and until 2018. This sample now includes a large majority of developing countries (N=106), with the exception of a few countries either in conflict or that stopped reporting data (e.g. Yemen, Iran). The bottom right panel replicates, for comparison, the main time series of effective tax rates for the unbalanced sample of developing countries that corresponds to Figure 2. Developing countries are defined in Section 3.3. This figure is discussed in Section 4.2.

Figure A4: Mechanism Impacts in Trade Liberalization Event Studies

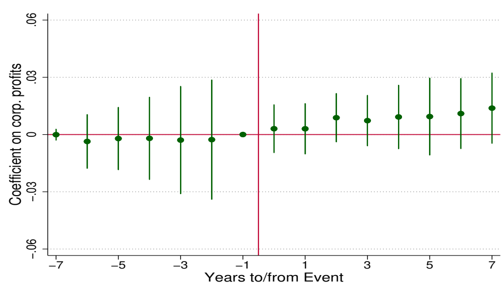
(a) Total taxes (% of NDP)



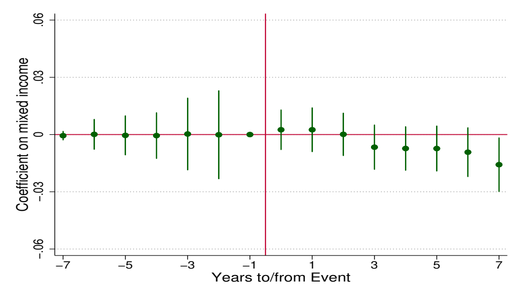

(b) Corp. income tax rate (first-diff.)



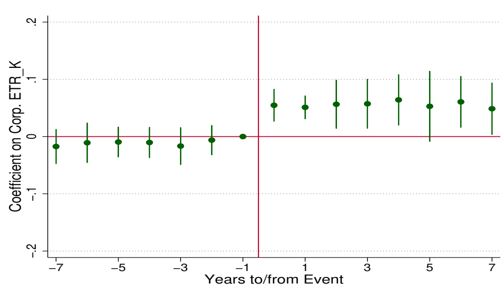

(c) Corporate income (% of NDP)



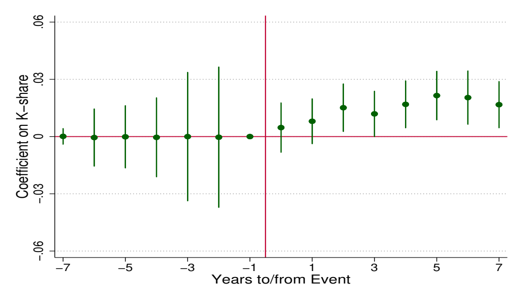

(d) Mixed income (% of NDP)




(e) Corporate $ETR_K$




(f) Capital income (% of NDP)




*Notes:* These panels show the impacts of the trade liberalization events on total taxes collected and mechanism outcomes. The panels are constructed using the method in Section 5.1, and similarly to Figure 6. Across panels, the outcome differs: panel a) is total tax revenue, as a percent of net domestic product (NDP); panel b) is the first-differenced statutory corporate income tax rate; panel c) is the corporate income share of net domestic product, where corporate income is the sum of corporate profits and employee compensation; panel d) is the mixed income share of net domestic product; panel e) is the average effective tax rate on corporate profits; panel f) is the capital share of net domestic product.

Table A1: Synthetic Difference-in-Difference of Trade Liberalization

| | Trade | $ETR_K$ | $ETR_L$ |
|---|---|---|---|
| | (1) | (2) | (3) |
| Panel A: *Synthetic control for each outcome separately* | | | |
| Post*Treat | 0.0551 | 0.0413*** | 0.0120 |
| | (0.0648) | (0.0147) | (0.00944) |
| Imputed treatment effect | 0.0551* | 0.0420*** | 0.0120*** |
| | (0.0292) | (0.00659) | (0.00498) |
| Panel B: *Synthetic control for all outcomes jointly* | | | |
| Post*Treat | 0.0697 | 0.0265 | 0.00293 |
| | (0.0429) | (0.0179) | (0.00632) |
| Imputed treatment effect | 0.0636* | 0.0249** | 0.00308 |
| | (0.0355) | (0.0125) | (0.00454) |
| Panel C: *Donor pool excluding major trading partners* | | | |
| Post*Treat | 0.0557 | 0.0412** | 0.0115 |
| | (0.0647) | (0.0145) | (0.00929) |
| Imputed treatment effect | 0.0557* | 0.0419*** | 0.0115** |
| | (0.0293) | (0.00652) | (0.00481) |
| Panel D: *Donor pool restricted to not-yet liberalized* | | | |
| Post*Treat | 0.0575 | 0.0426*** | 0.0112 |
| | (0.0628) | (0.0133) | (0.00903) |
| Imputed treatment effect | 0.0575* | 0.0431*** | 0.0112*** |
| | (0.0307) | (0.00852) | (0.00410) |
| Panel E: *Donor pool restricted to same region* | | | |
| Post*Treat | 0.0691 | 0.0415** | 0.00833 |
| | (0.0622) | (0.0187) | (0.00916) |
| Imputed treatment effect | 0.0691** | 0.0425*** | 0.00833* |
| | (0.0317) | (0.00716) | (0.00434) |
| Panel F: *Donor pool restricted to LMICs* | | | |
| Post*Treat | 0.0798 | 0.0415** | 0.0137 |
| | (0.0610) | (0.0148) | (0.00936) |
| Imputed treatment effect | 0.0798** | 0.0422*** | 0.0137*** |
| | (0.0331) | (0.00733) | (0.00442) |
| $N$ | 210 | 210 | 210 |

*Notes:* This table shows the results from estimating the difference-in-difference effect using the imputed treatment effect method - see Appendix C.2 for details. In Panel A, the synthetic control is created separately for each outcome and each liberalization country-event. In Panel B, the synthetic control is created for five outcomes jointly for each country-event: trade; $ETR_K$; $ETR_L$; the corporate share of net domestic product; the average effective tax rate on corporate profits. The event-study impacts on the last two variables are reported in Figure A4. In Panel C, the donor pool for each country-event excludes the 5 major import and export trading partners of the country, measured in terms of total volume of trade in the year immediately preceding liberalization. In Panel D, the donor pool excludes all countries that have already liberalized by the time of the event (based on Wacziarg & Welch, 2008). In Panel E, the donor pool is restricted to countries in the same region. In Panel F, the donor pool is all low and middle-income countries (LMICs), defined in Section 3.3.* p<0.10 ** p<0.05 *** p<0.01.

Table A2: First-Stage and Reduced Form Regressions

| | $1^{st}$-stage | Reduced form | | $1^{st}$-stage | | Reduced form | |
|---|---|---|---|---|---|---|---|
| | Trade<br>(1) | $ETR_K$<br>(2) | $ETR_L$<br>(3) | Trade<br>(4) | Trade$*\mathbb{1}$(High-inc.)<br>(5) | $ETR_K$<br>(6) | $ETR_L$<br>(7) |
| $Z^{gravity}$ | 0.068***<br>(0.010) | 0.010***<br>(0.003) | 0.003***<br>(0.001) | 0.012<br>(0.020) | -0.040***<br>(0.015) | 0.015<br>(0.009) | 0.008*<br>(0.004) |
| $Z^{oil-distance}$ | -0.110***<br>(0.034) | -0.034***<br>(0.010) | -0.020***<br>(0.006) | -0.087***<br>(0.018) | -0.026<br>(0.016) | -0.023***<br>(0.007) | -0.018***<br>(0.004) |
| $1^{st}$-stage F-statistic | 23.02 | | | 22.83 | 10.70 | | |
| $1^{st}$-stage Sanderson-Windmeijer Weak Instruments F-statistic | 23.02 | | | 41.54 | 23.71 | | |
| $1^{st}$-stage Kleibergen-Paap F-statistic | 23.02 | | | 14.68 | | | |
| Sample | Developing countries only | | | Developing and developed countries | | | |
| *N* | 4862 | 4862 | 4862 | 6489 | 6489 | 6489 | 6489 |

*Notes:* This regression table shows the first stage and the reduced form results. The sample is developing countries ($N = 4862$) in cols. (1)-(3), and developing and developed countries ($N = 6489$) in columns (4)-(7). Trade is exports and imports divided by net domestic product. Column (1) corresponds to the first-stage in developing countries, used in Tables 1-2-3. Columns (4)-(5) correspond to the first-stage in the full sample, which estimates heterogeneous effects by development level, and which is used in Table A7. We report several $1^{st}$-stage statistics: the F-statistic of excluded instruments; the Sanderson-Windmeijer multivariate F-test of excluded instruments; and, the Kleibergen-Paap F-statistic. When there is only one endogenous regressor (column 1), these three F-statistics are equivalent. Note in columns (4)-(5) that there is only one Kleibergen-Paap F-statistic, which evaluates the overall strength of the first-stage, even though there are two first-stage regressions. Columns (2)-(3) and (6)-(7) report the reduced form regressions of the instruments on the effective tax rates for capital, $ETR_K$, and labor, $ETR_L$. Developing (developed) countries are defined in Section 3.3. * p<0.10 ** p<0.05 *** p<0.01. Standard errors in parentheses are clustered at the country level.

## Table A3: Trade Impacts on Effective Tax Rates in Different Samples

| | Sample changes related to tax revenue data | | | | Sample changes related to System National Accounts data | | | Sample changes related to time-periods and balancedness | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Panel A: $ETR_K$ | | | | | | | | | | |
| Trade | 0.156*** | 0.178** | 0.256** | 0.297*** | 0.132* | 0.148** | 0.162*** | 0.154** | 0.265* | 0.151** |
| | (0.057) | (0.075) | (0.117) | (0.099) | (0.070) | (0.062) | (0.053) | (0.075) | (0.149) | (0.056) |
| Panel B: $ETR_L$ | | | | | | | | | | |
| Trade | 0.053*** | 0.028** | 0.093* | -0.027 | 0.057** | 0.054** | 0.035** | 0.043*** | 0.054 | 0.070*** |
| | (0.019) | (0.012) | (0.050) | (0.057) | (0.027) | (0.022) | (0.014) | (0.016) | (0.049) | (0.023) |
| Modifications to benchmark sample in Table 1 | Remove interpolated tax revenue | Only use HA tax data | Only use ICTD tax data | Only use OECD tax data | Remove composite SNA data | Remove SNA1968 data | Remove SNA2008 data | Only use pre-1994 years | Only use post-1994 years | Fully balanced panel 1965-2018 |
| $N$ | 4509 | 2214 | 1184 | 1297 | 2166 | 3932 | 3616 | 2086 | 2769 | 2425 |

*Notes:* This table presents results from estimating the effect of trade on effective tax rates in different samples across developing countries. The estimation is identical to the benchmark IV model in column (2) of Table 1; across columns, the sample differs from that benchmark sample. Developing countries are defined in Section 3.3. The outcome is the effective tax rate on capital, $ETR_K$, in Panel A and the effective tax rate on labor, $ETR_L$, in Panel B. Trade is measured as the sum of exports and imports divided by net domestic product (NDP). In the first four columns, sample-changes are made to the tax revenue data: interpolated values are dropped in column 1; the only data-source is historical archives (HA) in column 2; the only data-source is ICTD in column 3; the only data-source is OECD in column 4. In the next three columns, sample-changes are made to the system of national accounts (SNA) data: in column (5), the composite SNA values are removed; in column (6), data from SNA1968 is removed; in column (7), data from SNA2008 is removed. In the final three columns, sample-changes are made regarding balancedness: in column (8), the quasi-panel between 1965 and 1993 is used; in column (9), the quasi-panel between 1994 and 2018 is used; in column (10), the fully balanced panel of countries between 1965 and 2018 is used. For more details on the interpolations, imputations and data-sources, see Section 3 and Appendix B.

Table A4: Robustness of Results for Total Taxes and Mechanisms

| | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| Panel A: Total taxes (% of NDP) | | | | | | |
| Trade | 0.095 | 0.092** | 0.102*** | 0.112*** | 0.104*** | 0.190** |
| | (0.061) | (0.039) | (0.033) | (0.033) | (0.034) | (0.083) |
| *N* | 4862 | 3938 | 4862 | 4862 | 4862 | 4862 |
| Panel B: CIT rate (first-diff.) | | | | | | |
| Trade | 0.008 | -0.008 | -0.014** | -0.015*** | -0.014*** | -0.042** |
| | (0.012) | (0.007) | (0.005) | (0.005) | (0.005) | (0.019) |
| *N* | 4862 | 3938 | 4862 | 4862 | 4862 | 4862 |
| Panel C: $log(1{+}\text{CIT rate})$ | | | | | | |
| Trade | 0.005 | -0.006 | -0.010** | -0.011*** | -0.010*** | -0.033** |
| | (0.009) | (0.005) | (0.004) | (0.004) | (0.004) | (0.015) |
| *N* | 4862 | 3938 | 4862 | 4862 | 4862 | 4862 |
| Panel D: Corp. income (% of NDP) | | | | | | |
| Trade | 0.178*** | 0.197*** | 0.170*** | 0.187*** | 0.176*** | 0.221* |
| | (0.049) | (0.048) | (0.046) | (0.048) | (0.047) | (0.115) |
| *N* | 4862 | 3938 | 4862 | 4862 | 4862 | 4862 |
| Panel E: Mixed income (% of NDP) | | | | | | |
| Trade | -0.192*** | -0.162*** | -0.179*** | -0.190*** | -0.179*** | -0.140 |
| | (0.047) | (0.040) | (0.041) | (0.039) | (0.042) | (0.138) |
| *N* | 4862 | 3938 | 4862 | 4862 | 4862 | 4862 |
| Panel F: Capital share of NDP | | | | | | |
| Trade | 0.095* | 0.112** | 0.143*** | 0.157*** | 0.151*** | -0.021 |
| | (0.051) | (0.045) | (0.034) | (0.035) | (0.036) | (0.126) |
| *N* | 4862 | 3938 | 4862 | 4862 | 4862 | 4862 |
| Panel G: Corp. $ETR_K$ | | | | | | |
| Trade | 0.264* | 0.190* | 0.153** | 0.178** | 0.164** | 0.405** |
| | (0.149) | (0.096) | (0.079) | (0.080) | (0.079) | (0.195) |
| *N* | 4862 | 3938 | 4862 | 4862 | 4862 | 4862 |
| Modifications to IV in Panel B of Table 3 | NDP weights | Include country-year controls | Include **1**(oil-rich)*year fixed effects | Winsorize trade at 5%-95% | Only use $Z^{gravity}$ instrument | Only use $Z^{Oil-Dist}$ instrument |

*Notes:* This table presents robustness for trade's impacts on several outcomes in developing countries. Developing countries are defined in Section 3.3. Trade is the sum of exports and imports divided by net domestic product (NDP). The outcome differs across panels, and the specification differs across columns: each cell is the coefficient from a separate IV regression. The estimations are identical to the specifications in Table 1. Panel A is total taxes as a % of NDP. Panel B is the first-differenced corporate income tax (CIT) rate. Panel C is the percent change from log of $(1 + \text{CIT rate})$. Panel D is the corporate income share of NDP. Panel E is the mixed income share of NDP. Panel F is the capital share of NDP. Panel G is the average effective tax rate on corporate profits. * p<0.10 ** p<0.05 *** p<0.01. Standard errors in parentheses are clustered at the country level.

Table A5: Impacts of Trade in LMICs, Heterogeneity by Enforcement Policy

| | $ETR_K$ (1) | $ETR_L$ (2) | Corp. income (3) | Corp. $ETR_K$ (4) |
|---|---|---|---|---|
| Panel A: Large Taxpayer Unit | | | | |
| Trade | 0.113 | 0.015 | 0.169*** | 0.129 |
| | (0.069) | (0.031) | (0.057) | (0.106) |
| Trade∗𝟙(LTU) | 0.099 | 0.088** | 0.019 | 0.108 |
| | (0.079) | (0.040) | (0.051) | (0.122) |
| Implied coef. for | 0.212*** | 0.103*** | 0.188*** | 0.236** |
| Trade with LTU | (0.066) | (0.030) | (0.045) | (0.098) |
| Panel B: Customs-Tax Integration | | | | |
| Trade | 0.109 | 0.012 | 0.167*** | 0.118 |
| | (0.071) | (0.042) | (0.055) | (0.113) |
| Trade∗𝟙(Customs-Tax) | 0.290 | 0.257 | 0.066 | 0.351 |
| | (0.281) | (0.191) | (0.147) | (0.428) |
| Implied coef. for | 0.399 | 0.269 | 0.233* | 0.468 |
| Trade with Customs-Tax | (0.241) | (0.166) | (0.119) | (0.360) |
| Panel C: Value-Added Tax | | | | |
| Trade | 0.117** | 0.019 | 0.170*** | 0.130 |
| | (0.058) | (0.025) | (0.053) | (0.096) |
| Trade∗𝟙(VAT) | 0.111 | 0.099** | 0.023 | 0.123 |
| | (0.082) | (0.043) | (0.054) | (0.126) |
| Implied coef. for | 0.227*** | 0.117*** | 0.192*** | 0.252** |
| Trade with VAT | (0.068) | (0.033) | (0.049) | (0.102) |
| Panel D: International Accounting Standards | | | | |
| Trade | 0.130** | 0.023 | 0.157*** | 0.156* |
| | (0.057) | (0.023) | (0.052) | (0.093) |
| Trade∗𝟙(IAS) | 0.136 | 0.117*** | 0.016 | 0.162 |
| | (0.090) | (0.043) | (0.056) | (0.152) |
| Implied coef. for | 0.266*** | 0.140*** | 0.172*** | 0.317** |
| Trade with IAS | (0.080) | (0.038) | (0.054) | (0.130) |
| $N$ | 4862 | 4862 | 4862 | 4862 |

*Notes:* This table estimates heterogeneous IV effects of trade in developing countries (defined in Section 3.3.). Trade is the sum of exports and imports divided by net domestic product (NDP). Outcomes differ across columns: column (1) is the effective tax rate on capital, $ETR_K$; column (2) is the effective tax rate on labor, $ETR_L$; column (3) is the corporate income share of NDP; column (4) is the average effective tax rate on corporate profits. We estimate

$$y_{ct} = \mu \cdot trade_{ct} + \kappa \cdot trade_{ct} \cdot \mathbb{1}(A)_{ct} + \theta \cdot \mathbb{1}(A)_{ct} + \pi_c + \pi_t + \epsilon_{ct}$$

where $\mathbb{1}(A)_{ct}$ is an indicator variable which takes a value of 1 in all years after the administrative reform has been implemented. We instrument for $trade_{ct}$ and $trade_{ct} \cdot \mathbb{1}(A)_{ct}$ using the two instruments (Section 5.2). The coefficient on $\mathbb{1}(A)_{ct}$ is also estimated, but is not reported in the table. In Panel A, the administrative reform is the existence of a large taxpayer unit (LTU); this variable is coded based on the USAID's 'Collecting Taxes Database' (website link) and country-sources. In Panel B, the administrative reform is the integration of the customs authority and the domestic tax authority in a single revenue agency; this variable is coded based on USAID's 'Collecting Taxes Database' (website link), the OECD Tax Administration Comparative Series (website link), and country-sources. In Panel C, the administrative reform is the implementation of a value-added tax (VAT); this variable is coded based on Keen and Lockwood (2010) and country-sources. In Panel D, the administrative reform is the adoption of international accounting standards (IAS); this variable is coded based on the IAS country-profiles (website link). At the bottom of each column and panel, we report the implied coefficient and estimated standard error based on the linear combination of the $trade_{ct}$ and $trade_{ct} \cdot \mathbb{1}(A)_{ct}$ coefficients. * p<0.10 ** p<0.05 *** p<0.01. Standard errors in parentheses are clustered at the country level.

Table A6: Impacts of Trade Outside of Periods of Tax Revenue Loss

| | $ETR_K$ (1) | $ETR_L$ (2) | Corp. income (3) | Corp. $ETR_K$ (4) |
|---|---|---|---|---|
| Panel A: Excluding Trade-Induced Tariff Revenue Loss Periods (based on Cage and Gadenne, 2018) | | | | |
| Trade | 0.150** | 0.048** | 0.180*** | 0.181* |
| | (0.060) | (0.021) | (0.047) | (0.093) |
| $N$ | 3900 | 3900 | 3900 | 3900 |
| Panel B: Excluding Periods of Indirect Tax Revenue Loss | | | | |
| Trade | 0.192*** | 0.055*** | 0.196*** | 0.214** |
| | (0.054) | (0.017) | (0.047) | (0.088) |
| $N$ | 2977 | 2977 | 2977 | 2977 |
| Panel C: Excluding Periods of Total Tax Revenue Loss | | | | |
| Trade | 0.176*** | 0.049*** | 0.173*** | 0.197** |
| | (0.053) | (0.016) | (0.044) | (0.085) |
| $N$ | 2982 | 2982 | 2982 | 2982 |

*Notes:* This IV specification is the same as column (2) in Table 1, but modifications are made to the sample of developing countries. In Panel A, we exclude all country-year observations which belong to an episode of trade revenue loss, based on Cagé and Gadenne (2018). In a dataset of 130 countries between 1792 and 2006, the authors define such an episode by a fall in trade tax revenues as a percentage of GDP of at least 1 percentage point from a local yearly maximum to the next local yearly minimum that is accompanied by a non-decrease in the volume of imports as a share of GDP. In Panels B and C, we consider alternative definitions of revenue loss periods. In Panel B, we calculate the within-country yearly change in indirect taxes collected as a share of net domestic product (NDP), and take the three-year moving average. We then create terciles of this variable, separately for each country. We define periods of indirect tax revenue loss to be the observations which lie in the bottom tercile of this distribution, and exclude these country-year observations from the sample. In Panel C, we calculate the same revenue-loss variable, but based on changes in total taxes collected rather than indirect taxes collected. Trade is the sum of exports and imports divided by NDP. The outcome differs across columns: column (1) is the effective tax rate on capital, $ETR_K$; column (2) is the effective tax rate on labor, $ETR_L$; column (3) is the corporate income share of NDP; column (4) is the average effective tax rate on corporate profits. * p<0.10 ** p<0.05 *** p<0.01. Standard errors in parentheses are clustered at the country level.

Table A7: Heterogeneous Impacts of Trade by Development Level

| | $ETR_K$ | $ETR_L$ | First-diff. CIT Rate | Corp. Totl. Income | Mixed Income | Corp. Profits | Employee Comp. | Corp. $ETR_K$ | Natl. $K$-Share | Corp. $K$-Share |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Trade | 0.324** | 0.159* | -0.018 | 0.300** | -0.198* | 0.186*** | 0.065 | 0.457** | 0.139** | 0.168** |
| | (0.147) | (0.094) | (0.023) | (0.127) | (0.117) | (0.054) | (0.109) | (0.205) | (0.057) | (0.066) |
| Trade$*\mathbb{1}$(High-inc.) | -0.241 | 0.107 | -0.062* | -0.500*** | 0.342** | -0.318*** | -0.213* | -0.288 | -0.197** | -0.238** |
| | (0.237) | (0.263) | (0.032) | (0.152) | (0.136) | (0.096) | (0.113) | (0.319) | (0.087) | (0.110) |
| Implied coef. for Trade in High-inc. | 0.084 | 0.266 | -0.080*** | -0.199 | 0.144 | -0.132** | -0.148 | 0.170 | -0.057 | -0.070 |
| | (0.128) | (0.189) | (0.019) | (0.130) | (0.136) | (0.063) | (0.118) | (0.164) | (0.052) | (0.076) |
| $N$ | 6489 | 6489 | 6489 | 6489 | 6489 | 6489 | 6489 | 6489 | 6489 | 6489 |

*Notes:* This table presents IV results from estimating the effects of trade on $ETR$ and mechanism outcomes in the full sample of developing and developed countries (Section 3.3). Trade is measured as the sum of exports and imports divided by net domestic product (NDP). We run the following IV regression: $y_{ct} = \mu \cdot trade_{ct} + \kappa \cdot trade_{ct} \cdot \mathbb{1}(HighIncome)_c + \Theta \cdot X_{ct} + \pi_c + \pi_t + \epsilon_{ct}$ The first-stage regression is reported in Table A2. At the bottom of each column, we report the implied coefficient and estimated standard error based on the linear combination of the $Trade$ and the $Trade * \mathbb{1}$(High-inc.) coefficients. Each column is a different outcome: column (1) is the effective tax rate on capital (winsorized); column (2) is the effective tax rate on labor (winsorized); column (3) is the first-differenced statutory corporate income tax rate; column (4) is the corporate income share of net domestic product, where corporate income is the sum of corporate profits and corporate employee compensation; column (5) is the mixed income share of net domestic product; column (6) is the corporate profit share of net domestic product; column (7) is the employee compensation share of net domestic product; column (8) is the average effective tax rate on corporate profits; column (9) is the capital share of net domestic product; column (10) is the capital share of corporate income. In columns (1)-(2), the outcome is winsorized and the regression includes a set of interactions between year dummies and a dummy for the top decile of the outcome variable. These extensions help to control for the influence of large values of $ETR$ on the estimated coefficient of trade. For more details on outcomes, see Section 3.1 and Section 6.2. For more details on the instrumental variables, see Section 5.2. * p<0.10 ** p<0.05 *** p<0.01. Standard errors in parentheses are clustered at the country level.

# Appendix B Data & Construction of Effective Tax Rates

This appendix section provides an overview of the data sources used to create our tax revenue and national income series (Section B.1). Additionally, we discuss the methodology to measure effective tax rates (Section B.2).

## B.1 Data sources

**Tax revenue data** Our tax revenue data draws from three key sources:

(i) **OECD Government Revenue Statistics** (website link): OECD revenue statistics take precedence in our data hierarchy as it contains all types of tax revenues already arranged in the OECD taxonomy of taxes. While it covers all OECD countries, it only covers a subset of developing countries which typically start in the early 2000s.

(ii) **ICTD Government Revenue Dataset** (website link): ICTD data covers many developing countries, but only begins in the 1980s. ICTD at times does not separate income taxes into personal vs. corporate taxes and often does not contain social security contributions.

(iii) **Archival data**: The main archival data collection corresponds to the digitization of the Government Documents section in the Lamont Library at Harvard University (website link). For each country, we scanned, tabulated and harmonized official data from the public budget and national statistical yearbooks, to retrieve official tax revenue statistics. The supplementary appendix lists the main historical documents used in each country's time-series. In the case where the document is a statistical yearbook, the initial listed source is always a report produced by the finance ministry or the national tax authority. To complement hard-copy archival data, we retrieved countries' online reports, usually published by their national statistical office or finance ministry. We also used complementary sources, including offline archival Government Finance Statistics data from the IMF which covers the period 1972-1989. For social security contributions, we relied on two additional sources: the 'D61' statistic on social contributions in the household sector in SNA-1968 and SNA-2008, and data from Fisunoglu, Kang, Arbetman-Rabinowitz, and Kugler (2011).

To increase the credibility of the tax revenue series based on newly digitized historical documents, we base our approach on the following four guiding rules:

1. We seek to build long time-series from the archival records in order to overlap with pre-existing sources (OECD, ICTD, IMF). We use the overlapping years to inspect that the different sources provide similar estimates of the overall levels of taxes collected and to verify that they report the same set of taxes in place. If discrepancies exist when data sources overlap, we inspect the accuracy of each source with additional information. For this reason, switches in data-source rarely lead to a significant change in trend.

2. In historical time-periods where no overlap exists with pre-existing sources, we find academic publications and policy reports to compare the estimated overall levels of

tax/GDP. When discrepancies exist, we investigate its causes (e.g. inclusion of non-tax revenues, differences in estimated GDP numbers).

3. We take note of instances where the overall tax take, or individual tax types, see sudden and large changes. We use additional sources to try to determine the proximate causes as they relate to policy changes, political transitions or economic shocks. We flag cases where we cannot find the proximate cause or where the political or economic events induce very significant volatility in the time-series.

4. We aim to be conservative in our inclusion of countries and time-periods. Specifically, we exclude countries in time-periods where data exists but where significant concerns remain about its reliability (and where it proves difficult to find corroborating sources). These instances are often in periods of significant political or economic change. For example, we exclude Afghanistan in the late 1970s and early 1980s; Cambodia in the late 1980s and early 1990s; Dominican Republic in the early 1960s; and, Namibia in 1990.

The supplementary appendix contains a table which summarizes our decisions as they relate to these four guiding rules in each country in our sample. The table emphasizes the uncertainty that exists for specific countries in specific time periods and we flag instances where we assess the data to be worthy of inclusion but where it should still be interpreted with caution and additional investigations would be helpful. We confirm that none of our main results change if we exclude these flagged instances. Moreover, the supplementary appendix provides case-studies with additional details on our decisions and direct links to the initial historical documents for each country. The case studies are currently limited to 67 countries but will ultimately cover the entire sample.[66]

Equipped with the historical time series, we have to construct long-run panels across sources. Below, we outline the instructions used to harmonize across sources and to improve data quality for the measurement of each type of tax. We flag instances where we consider the series to be legitimate, but where harmonization proved more challenging due to coinciding economic or political changes. For each country, the main decisions related to harmonization and data-quality are provided in the supplementary appendix.

1. We first rely on OECD data whenever it exists. Archival data is initially second in priority, but we revise this based on whether ICTD data provides a long time series and separates personal from corporate income taxes. We also study if ICTD has the better match in overlapping time-periods with OECD data. When possible, we aim to use no more than two data sources per country.

2. We exclude country-years for communist/command economies. This implies that our panel size jumps in 1994, including when China and Russia first appear. The year 1994 is a few years removed from the dissolution of the Soviet Union but, as discussed below, arguably corresponds to China's establishment of a modern tax system (World Bank, 2008).

[66]We invite comments from researchers to improve the accuracy of the series as we build the case studies and expand the data to recent years.

3. When none of the data sources separate PIT from CIT, we use academic sources and tax legislation to assign values.

4. To guard against omitting significant values of decentralized tax revenues, we use the OECD database on subnational government finance (link) to find the countries with significant state and local taxes, and we attempt to collect further data for these countries if necessary.

5. We linearly interpolate data when a given tax type is missing, but for no more than 4 years in a time-series and without extrapolation. We check for significant socio-economic changes that could cast doubt on the continuity of the tax revenue series and do not interpolate in such years.

6. We only use actual amounts of taxes collected, and do not rely on estimated values.

**China's establishment of a modern tax system in 1994**

In our benchmark setting, we only include formerly communist economies into our data starting in 1994. Given China's weight in the global economy, it is worth reviewing the reason for that choice. The tax revenue data for China covers most of our sample period although its quality improves markedly in the 1980s. Official statistics are available online: link here.

Prior to the 1980s, China had a command economy model of 'profit delivery,' in which the state directly received the revenues of profitable SOEs, and subsidized unprofitable ones. A corporate income tax first appears in China in 1983-84, but the majority of the base continues to be state-owned enterprises. In 1985, the tax system was further reformed into a 'fiscal contracting' system whereby firms negotiated a fixed lump-sum payment (regardless of economic outcomes), which cannot be split into labor versus capital taxes (nor into consumption taxes). We therefore exclude the 'pseudo'-CIT revenue dating from 1985 through 1993.

Rather, we consider that China's modern tax system began in 1994. The World Bank (2008) shows that, in 1994, China established for the first time a central tax administration; reformed the 'fiscal contracting' system; unified the PIT; created a VAT; and reduced 'extra budgetary' (non-tax) revenues. Thus from 1994 onward we can categorize tax revenue precisely by type, assign them to capital or labor, and estimate our $ETR$s.

**National accounts data** To compute factor incomes of net domestic product, we combine two main datasets from the United Nations Statistics Division. The first is the 2008 System of National Accounts (SNA) online data repository. The second is the 1968 SNA archival material. The 2008 and 1968 SNAs initially have different reporting classifications; to the best of our knowledge, our project is the first to harmonize national accounts across these two sources.

To estimate capital and labor factor incomes requires information on the 4 main sub-components that make up net domestic product (see equation 3). However, in some country-years where we have information on domestic product from an SNA dataset, there may not be data on all four sub-components at the same time. This is more frequently the case for the 1968 SNA than for the 2008 SNA and it is most frequent for mixed income ($OS_{PUE}$). In these cases, we first attempt to recover the value of the missing component using data from the other SNA dataset and national accounting identities with non-missing values for other components within the same country-year. For the remaining cases after applying this process, we impute values for the component. All of the regressions in Sections 5-6 include dummy variables for these composite cases; our main results also hold without the imputed values (Table A3). For the imputation, we follow the procedure from Blanchet et al. (2021). The World Inequality Database uses this procedure to impute consumption of fixed capital (depreciation) when it is missing in countries' series. For example, applying this procedure in our setting means that we model $OS_{PUE}$ as a function of log national income per capita, a fixed country characteristic, and an AR(1) persistence term.

Table B1 summarizes the national accounts coverage in our dataset. The 'Complete SNA2008' row refers to country-years where all components of net domestic product are extracted from the 2008 SNA; similarly for the 'Complete SNA1968' row. The 'Composite' row counts instances where one component (or more) of net domestic product is initially missing from an SNA dataset and is retrieved from the other SNA dataset, is calculated via accounting identities, or is imputed.

Table B1: Main Data Sources

| | Country-year obs. | % |
|---|---|---|
| Panel A: Tax revenue data | | |
| OECD | 2879 | 42.34% |
| Archives | 2657 | 39.08% |
| ICTD | 1263 | 18.58% |
| *N* | *6799* | *100%* |
| Panel B: Factor income data | | |
| Complete SNA2008 | 2448 | 36.01% |
| Complete SNA1968 | 1398 | 20.56% |
| Composite | 2953 | 43.43% |
| *N* | *6799* | *100%* |

*Notes:* See Section B.1 for more details on the data-sources for tax revenue and factor income.

## B.2 Construction of $ETR$

By combining data on disaggregated tax revenues and national income components, we construct effective tax rates on capital and labor (equations 1 and 2 in Section 3.1). Here we provide further details on the definitions of $ETR$. Computing $ETR_L$ and $ETR_K$ requires the following information for country $c$ in year $t$:

$$ETR_{L,ct} = \frac{T_{L,ct}}{Y_{L,ct}} = \frac{\lambda_{PIT,ct} \cdot T_{1100,ct} + \lambda_{socsec,ct} \cdot T_{2000,ct}}{CE_{ct} + \phi_{ct} \cdot OS_{PUE,ct}}$$

$$ETR_{K,ct} = \frac{T_{K,ct}}{Y_{K,ct}} = \frac{(1-\lambda_{PIT,ct}) \cdot T_{1100,ct} + (1-\lambda_{CIT,ct}) \cdot T_{1200,ct} + (1-\lambda_{assets,ct}) \cdot T_{4000,ct}}{(1-\phi_{ct}) \cdot OS_{PUE,ct} + OS_{CORP,ct} + OS_{HH,ct}}$$

For each type of tax $j$, there is a $\lambda_{j,ct}$ allocation of the tax to labor which may vary by country-year (and $1-\lambda_{j,ct}$ is the allocation to capital). The allocation for each type of tax is described in Table B2, where the types of taxes follow the OECD classification. In our benchmark assignment, these allocations are time- and country-invariant for all types of taxes, except for personal income taxes ($\lambda_{PIT,ct}$) which we discuss in detail below. Further, in our benchmark assumption, we assume that the labor share of mixed income, $\phi_{ct}$, is fixed at 75% in all country-years ($\phi_{ct} = 0.75$). In robustness checks, we let $\phi_{ct}$ vary at the country-level, based on ILO (2019), or at the country-year level by using the labor share in the corporate sector. In our benchmark assignment, replacing the invariant parameters with their fixed numerical values, we therefore have:

$$ETR_{L,ct} = \frac{T_{L,ct}}{Y_{L,ct}} = \frac{\lambda_{PIT,ct} \cdot T_{1100,ct} + T_{2000,ct}}{CE_{ct} + 0.75 \cdot OS_{PUE,ct}}$$

$$ETR_{K,ct} = \frac{T_{K,ct}}{Y_{K,ct}} = \frac{(1-\lambda_{PIT,ct}) \cdot T_{1100,ct} + T_{1200,ct} + T_{4000,ct}}{0.25 \cdot OS_{PUE,ct} + OS_{CORP,ct} + OS_{HH,ct}}$$

The parameter values are described in Table B2, both for the tax revenue numerator and the national income denominator. We now provide more details on $\lambda_{PIT}$ and $\phi$.

**Labor share of personal income taxes:** $\lambda_{PIT}$ As discussed in Section 3.1, the level of personal income tax (PIT) that derives from capital versus labor income is rarely directly observed.[67] Thus, within PIT, an important parameter is the share of revenue assigned to labor, denoted $\lambda_{PIT}$. In the United States, Piketty et al. (2018) find that approximately 85% of PIT revenue is from labor and 15% from capital. To construct country-year specific $\lambda_{PIT,ct}$, we start from the US benchmark ($\lambda_{PIT} = 85\%$) and make two adjustments:

(a) First, the location of the PIT exemption threshold in the income distribution impacts $\lambda_{PIT}$, since the capital income share is higher for richer individuals. We retrieve PIT exemption thresholds from Jensen (2022). We assume countries with a higher PIT exemption threshold have a higher $\lambda_{PIT}$. Since the US has a low exemption threshold

[67] PIT revenue from capital income includes taxes on dividends and capital gains and on the capital share of self-employment income. OECD revenue data occasionally reports tax revenue from capital gains, which was on average 4% of PIT in the period 2010-2018 (7.5% in the US).

with $\lambda_{PIT} = 85\%$, we assign 85% of PIT to labor in countries where the PIT at least half of the workforce (mainly high-income countries). For countries where the PIT covers 1% or less of the workforce (lowest-income countries), we assign a maximum PIT capital share of 30%. For PIT thresholds with a coverage between 1% to 50% of the workforce, we linearly assign $\lambda_{PIT}$ between 70% and 85%.

(b) Second, we assume that countries where a dual PIT system is in place have a larger $\lambda_{PIT}$. Dual PIT systems set capital income taxation to a lower—often flat—rate, while labor income is taxed with progressive marginal tax rates. We compute the measure of the percent difference between the tax rate on dividends and the top marginal tax rate on labor income. Data on dividend vs wage income tax rates are taken from OECD Revenue Statistics and country-specific tax code documents. Since we only have dividend rates, we assume that 50% of capital income in PIT benefits from the lower rate (e.g., capital gains might not benefit). For this 50%, we multiply $\lambda_{PIT}$ by the percent difference in dividend versus top marginal tax rates.

**Labor share of mixed income:** $\phi$ Section 3.1 noted the difficulty of estimating the labor share of mixed income (unincorporated enterprises). We assume a benchmark measure of $\phi = 75\%$. The implied capital share is lower than the 30% used in Distributional National Accounts guidelines (Blanchet et al., 2021). However, since the global average corporate capital share is 27%, assuming that the capital share of unincorporated enterprises is slightly lower appears reasonable (see Guerriero, 2019).

We implement two robustness checks. First, we set the labor share of mixed income equal to that of the corporate sector at the country-year level; specifically, $\phi_{ct} = \frac{CE_{ct}}{CE_{ct}+OS_{CORP,ct}}$. This procedure follows Gollin (2002).

Second, we implement the ILO (2019) method which relies on harmonized household surveys and labor force surveys in developing countries between 2004 and 2017. Estimation of the relative labor income of self-employed is based on the observable characteristics of those workers and their comparison with employees. Relevant variables, including industry, occupation, education level and age, are used in a regression to uncover the determinants of labor income of employees. Given the estimated relationship between employee labor income and the explanatory variables, labor income is extrapolated to self-employed, generating a coefficient of relative earnings to employees, denoted $\gamma_q$. The method estimates a separate $\gamma_q$ for different groups $q$ of self-employed: self-employed workers; own-account workers; and, contributing family members. A correction procedure is implemented to reduce the bias from selection into self-employment. Total labor income in a given country-year is then determined as $Y_L^{ILO} = CE + \sum_q w_{emp} \cdot \gamma_q \cdot b_q$, where $CE$ is the total compensation of employees in SNA, $w_{emp}$ is the average employee wage (which relates $CE$ to the total employee workforce), $b_q$ is self-employed group $q$'s count in the workforce, and $\gamma_q$ is the $q$-specific earnings coefficient relative to the average employee wage. Equipped with the $Y_L^{ILO}$ estimate, we calculate the 'implicit' labor mixed income ($OSPUE_L$) as the difference between $Y_L^{ILO}$ and the value of compensation for employees $CE$ observed in the national accounts. Then, we compute the mixed income share allocated to labor. Specifically, $\phi^{ILO}$ is computed as follows: $\phi^{ILO} = \frac{(Y_L^{ILO} - CE)}{OSPUE} = \frac{OSPUE_L^{ILO}}{OSPUE}$

Finally, we compute the average $\phi^{ILO}$ for each country during 2004-2017 and assign this value to all years. We assign a country-specific but time-invariant value for two reasons. First, prior to 2004, the ILO lacks the required data to compute $Y_L^{ILO}$ on a country-year basis. Second, when measured at the country-year level during the 2004-2017 period, $\phi^{ILO}$ varies little within country across years. Assigning a country-specific but time-invariant mixed income factor share may therefore be reasonable.

The main challenge is that the estimation framework for $\gamma_q$ is not disciplined by the country's actual values in SNA. In particular, nothing prevents $\sum_q w_{emp} \cdot \gamma_q \cdot b_q > OS_{PUE}$ - such that estimated labor mixed income is larger than the SNA actually observed entire mixed income. This would, implausibly, imply that $\phi^{ILO} > 100\%$. To remedy this concern, we winsorize $\phi^{ILO}$ at 100%. In cases where $\gamma_q$ and $b_q$ are not from ILO (2019), we also winsorize $\phi^{ILO}$ from below with the lowest observed country value in ILO (2019), which is 36%. While the ILO (2019) method generates important country-level variation, the global average value for $\phi^{ILO}$, at 80%, does not differ much from our benchmark value $\phi = 75\%$.

**Mixed income in China and the US** We make mixed-income adjustment to the benchmark series for China and the United States. For China, Piketty, Yang, and Zucman (2019) (PYZ) show that Chinese national accounts systematically underestimate mixed income and overestimate other factor incomes: for example, the income of self-employed agricultural workers is attributed to employee compensation in the SNA 2008 data and not to mixed income (as in other countries). We base our mixed income series on PYZ.

Following PYZ, we define mixed income as the sum of the income attributed to self-employed workers from agriculture and individual businesses. PYZ covers the period 1992-2014. For years before and after, we extend the series as follows:

(a) For agriculture, relevant data is available dating back to 1952. We extend the series back to 1965 relying on the price deflator available at World Inequality Database. For more recent years (2014-2018), we predict the trend based on sources used in PYZ (National Bureau of Statistics, link).

(b) For individual businesses, PYZ computes the income of this sector by combining several data sources. Unfortunately, a crucial part of it is not available prior to 1992, namely the 'flow of funds' data. Instead, our assumption is that, prior to 1980, Chinese individual businesses accounted for a negligible share of the economy. This observation is consistent with facts on self-employment structure in China at the micro and macro levels, and the trends presented in PYZ for the 1990s.[68] For recent years (2014-2018), we predict the trend based on sources used in PYZ (National Bureau of Statistics, link).

The estimated series of mixed income in China follows the same trend as for the rest of LMICs, although it starts from a slightly higher initial level.

For the US, we use the factor shares from Piketty et al. (2018), which (i) assumes a higher capital share of income for partnerships vs. other non-corporate businesses; and (ii) accounts for the rising capital intensity of partnerships since the 1980s.

[68] At the micro level, self-employed workers represent less than 2% of workforce in the 1980s, but had similar income per capita as wage earners (Gustafsson & Zhang, 2022). At the macro level, very small-scale industries represented 0.4% of industry output in the 1970s, reaching 7% only in 1989 (Yusuf, 1994).

## Table B2: Main Tax Revenue and National Accounts Concepts

*Panel A: Tax Revenue*

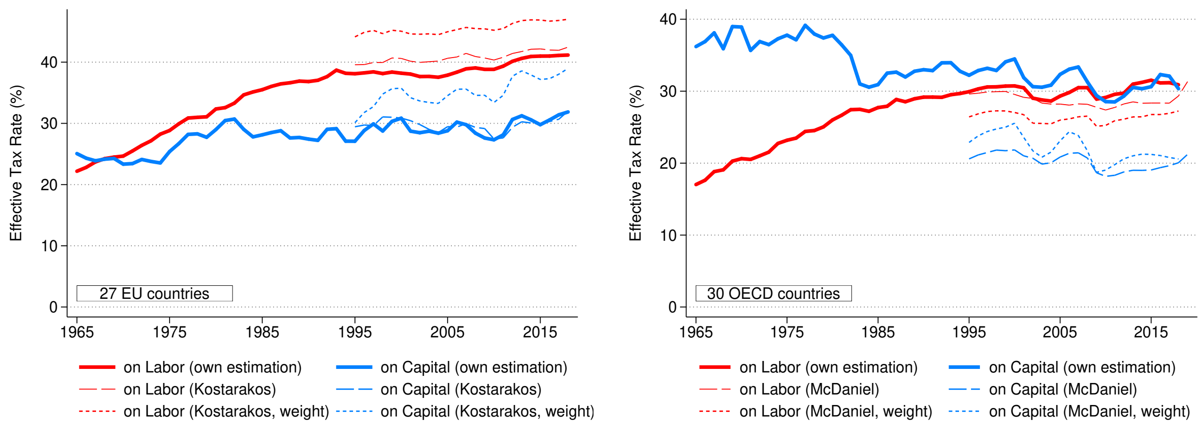
| OECD revenue classification | **type of tax $j$** | **incidence $\lambda_j$ on labor** | **notes** |
|---|---|---|---|
| *1100* | personal income tax (PIT) | $68\% \leq \lambda_{PIT} \leq 93\%$ | Taxes on individuals (wages, capital income, capital gains). $\lambda_{PIT,ct}$ varies by country and year: see Section B.2 for details |
| *1200* | corporate income tax (CIT) | $\lambda_{CIT} = 0\%$ | Taxes on corporate profits. Unallocable income taxes (OECD category *1300*) are split between PIT and CIT based on information from additional sources (see supplementary appendix) |
| *2000 / 3000* | social security & payroll | $\lambda_{soc.sec.} = 100\%$ | Includes all social security contributions as well as payroll taxes |
| *4000* | property & wealth taxes | $\lambda_{assets} = 0\%$ | Includes property, wealth and financial transaction taxes |
| *5000* | indirect taxes | excluded | Includes trade taxes, value-added taxes and other sales taxes and excise taxes. We consider these taxes as prior to factor income returns, such that they can be excluded from factor income taxation (Browning, 1978; Saez and Zucman, 2019). |
| *6000* | other taxes | excluded | Rare in occurrence and often quantitatively small |
| *7000* | non-tax revenue | excluded | Does not meet definition of taxation, can be quantitatively significant |

*Panel B: National Accounts*

| Natl. accounts acronym | **national income component** | **benchmark allocation** | **notes** |
|---|---|---|---|
| $CE$ | compensation of employees | labor | Includes wages and salaries, employer and employee social contributions, and all payments from employers to their employees |
| $OS_{PUE}$ | mixed income | $\phi = 75\%$ labor | 'Operating surplus of private unincorporated enterprises' includes income from self-employment, household business owners, and informal or unincorporated enterprises |
| $OS_{HH}$ | imputed rent | capital | 'Operating surplus of households' is imputed rental income accruing to homeowners who live in their own home |
| $OS_{CORP}$ | corporate profits | capital | 'Operating surplus of corporations' includes all corporate income after paying employees and expenses, and can be thought of as corporate-sector capital income |
| $OS_{GOV}$ | government operating surplus | — | $OS_{GOV} = 0$, by construction in national accounts |
| $NIT$ | net indirect taxes | excluded | 'indirect taxes, net of subsidies' usually comprise 8-15% of national income. |
| $NFI$ | net foreign income | — | We treat domestic income without balancing the accounts to foreign earned income: many countries tax income earned domestically, regardless of citizenship, whereas net foreign income is taxed only with difficulty |
| $CFC$ | depreciation | excluded | Factor income and our $ETR$ are expressed net of 'consumption of fixed capital' |

Figure B1: Comparing $ETR$ Evolution in Our Data and Existing Studies

(a) Kostarakos and Varthalitis (2020) (b) McDaniel (2020)




*Notes:* These graphs provide a comparison of our $ETR$ estimations with the recent literature. The left-hand graph compares our estimations with Kostarakos and Varthalitis (2020), based on EU-27 members from 1995 to 2019. The right-hand graph compares our estimations with the updated dataset in McDaniel (2020) that includes 30 OECD countries from 1995 to 2018. This extension is based on McDaniel (2007) (Table B3), and covers the largest OECD countries, including the US, as well as Mexico and Turkey. The solid line represents the results using our $ETR$ measures and weights, but based on the exact country samples in the respective studies. The long-dash line replicates the $ETR$ measures from the two studies. The short-dash line extends their $ETR$ series but using our country-year weights. For a discussion of the differences between series, see Section 4.2, Table B3 and the supplementary appendix.

Table B3: Effective Tax Rates: Existing Databases

| Paper | Time | Countries | Source | Notes on methodological differences with our approach |
|---|---|---|---|---|
| Mendoza et al (1994) | 1965-1988 | G7 members | OECD | Difference: All mixed income is allocated to capital income.<br>Difference: Labor and capital in the PIT are taxed at the same rate |
| Carey and Rabesona (2004) | 1975-2000 | 25 OECD biggest members | OECD | Difference: Mixed income allocation where self-employed pay themselves the annual salary earned by the average employee.<br>Similarity: Labor and capital in PIT are not taxed at same rate, measure preferential tax treatment of pension funds and dividends.<br>Difference: Social security contributions deducted from household income. |
| McDaniel (2007) (McDaniel 2020) | 1950-2003 (updated: 1995-2018) | 15 OECD biggest members (updated 30 OECD biggest members) | OECD | Difference: Mixed income imputed to capital based on rest-of-economy share.<br>Difference: Labor and capital in PIT are taxed at the same rate |
| Kostarakos and Varthalitis (2020) | 1995-2019 | EU-27 members | Eurostat | Follows Carey and Rabesona (2004) |

# Appendix C  Trade Liberalization Event Studies

## C.1  Description of liberalization events

Our selection of trade events is determined by three criteria. First, the event is related to measurable policy reforms; this improves the transparency of the event-study design which is based on a well-defined policy event. Second, the policy reforms induced large changes in trade barriers; this increases the likelihood of observing sharp breaks in macroeconomic outcomes around the event-time. Third, the event has been studied in academic publications; this allows us to rely on events for which the positive effects on openness have previously been established.

These criteria led us to focus on the six trade liberalization events referenced in review articles by Goldberg and Pavcnik (2007, 2016) to which we add China's WTO accession event (studied in Brandt et al., 2017). Most of these selected events feature reductions in tariff rates: many of the countries did not participate in the early GATT/WTO negotiation rounds, making reductions in tariffs an available policy lever. The tariff reductions were large: Brazil cut tariff rates from 59% to 15%, India from 80% to 39% percent, and China from 48% to 20%. Mexico reduced tariff rates from 24% to 12% and import license requirements went from covering 93% of national production to 25%; Colombia's tariffs were reduced from 27% to 10% and import requirements dropped from 72% of national production coverage to 1%. In the selected countries, "tariff reductions constitute a big part of the globalization process" (Goldberg & Pavcnik, 2016). The timing of the events and academic references are provided in the supplementary appendix.

Below are narrative analyses for some of the events:

- **Brazil** The liberalization event of 1988 is detailed in Dix-Carneiro and Kovak (2017). The authors note: "In an effort to increase transparency in trade policy, the government reduced tariff redundancy by cutting nominal tariffs... Liberalization effectively began when the newly elected administration suddenly and unexpectedly abolished the list of suspended import licenses and removed nearly all special customs regimes."

- **Colombia** Similarly to Brazil, tariff reductions in Colombia in 1985 were driven by the country's decision to impose uniform rates across products and industries under the negotiation commitments to the WTO. Goldberg and Pavcnik (2007) note that this reform objective makes "the endogeneity of trade policy changes less pronounced here [in Colombia] than in other studies."

- **China** Brandt et al. (2017) note that trade openness reforms had gradually been implemented in China prior to the country's WTO accession in 2001, but that the tariff reductions implemented upon accession were large, "less voluntary" and largely complied with the pre-specified WTO accession agreements. Importantly, the potential accession to WTO contributed to the timing of privatization initiatives, in which the Chinese government restructured and reduced its ownership in state-owned enterprises. While the privatization efforts began in 1995 and were incremental, it is possible that additional sell-offs in the post-WTO years contribute to the observed medium-run trends in our outcomes.

- **India** The 1991 event in India occurred as a result of an IMF intervention that dictated the pace and scope of the liberalization reforms. Under the IMF program, tariff rates had to be harmonized across industries, which, like in Brazil and Colombia, led to a large average reduction in tariffs. Topalova and Khandelwal (2011) argue the Indian reform "came as a surprise" and "was unanticipated by firms in India." The reforms were implemented quickly "as a sort of shock therapy with little debate or analysis." The IMF program was in response to a set of events including "the drop in remittances from Indian workers in the Middle East, the increase in oil prices due to the Gulf War, and political uncertainty following the assassination of Rajiv Gandhi".

- **Vietnam** The 2001 reform was implemented as a broad trade agreement that did not involve negotiations over specific tariffs (McCaig & Pavcnik, 2018). The reform was driven by the American government's decision to reclassify Vietnam from 'Column 2' of the US tariff schedule to 'Normal Trade Relations'. Column 2 was designed in the early 1950s for the 21 communist countries, including Vietnam, with whom the US did not have normal trading relations.

These descriptions of reform timing do not suggest that the liberalization events were directly triggered by changes in domestic taxation or factor incomes.

Goldberg and Pavcnik (2007) note other cross-border reforms that occurred during post-years of the liberalization events. Argentina's 1989 event and Brazil's 1988 event were followed by accession to Mercosur in 1991; India's 1991 event was followed by foreign direct investment liberalization in 1993; and Mexico's 1985 WTO accession was followed by a removal of capital inflow restrictions in 1989. These reforms occurred with some lag to the trade liberalization events.

## C.2 Event study methodology

Our sample is constructed by applying a synthetic matching procedure to every treated country for each outcome of interest. The donor pool has to be fully balanced in all pre-event periods. To estimate the event study in equation (4) for a given outcome, the sample pools the seven treated countries and their synthetic control countries for 7 years before and after the events (yielding 210 observations). We estimate the event-study in equation 4 and the DiD model: $y_{ct} = \beta^{DiD} \cdot \mathbb{1}(e \geq 0)_t \cdot D_c + \theta_t + \kappa_c + \pi_{Year(t)} + \epsilon_{ct}$. The DiD model uses the same notation as equation (4). Moreover, we use the imputation method by Borusyak et al. (2021) to report average treatment effects comparable to $\beta^{DiD}$ with a technique that deals with issues with two-way fixed effects and heterogeneous event timing. Details are provided in the supplementary appendix. All the DiD average treatment effects are reported in Table A1.

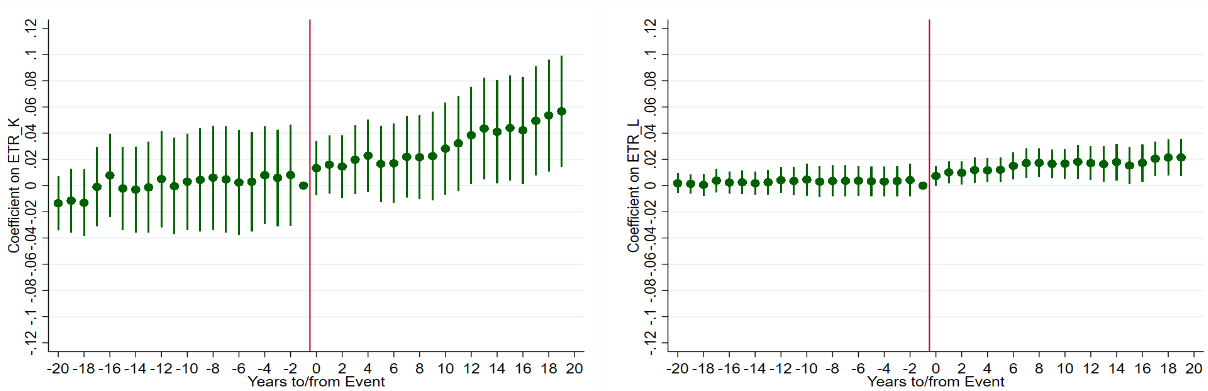
## C.3 Alternative trade liberalization event study

We present results based on an alternative measure of trade liberalization events. We use the events from Wacziarg and Welch (2008), which cover 141 countries at all levels of development between the 1950s and 1998. When merged with our data, the sample covers 68 liberalization events that occurred between 1965 and 1998 in developing countries. A trade liberalization event is defined to occur when all five of the following conditions no longer hold: (i) average tariff rates are above 40%; (ii) non-tariff barriers cover at least 40% of trade; (iii) the black market exchange rate is at least 20% lower than the official exchange rate; (iv) there is a state monopoly on major exports; (v) there is a socialistic system in place. These conditions are broader than our main liberalization event criteria (Section 5.1 and C.1). At the same time, our main events are covered in this expanded event sample (with the exception of China and Vietnam, whose events are after the end of the sample period); this occurs because the reduction in tariff rates, one of our main event criteria, was the remaining event-condition to be satisfied in Wacziarg and Welch (2008). We estimate the effects of the liberalization events using the DiD model: $y_{ct} = \beta^{DiD} \cdot E_{ct} + \theta_t + \theta_c + \epsilon_{ct}$. $y_{ct}$ is the outcome of interest in country $c$ in year $t$, $E_{ct}$ is the event indicator which takes on a value of 1 in all periods after a country has a liberalization event (and 0 otherwise), and $\theta_t$ and $\theta_c$ are year and country fixed effects, respectively. $\epsilon_{ct}$ is clustered at the country level. Estimation issues arising from heterogeneous treatment-timing may be important; for this reason, we focus on the imputed treatment effects based on Borusyak et al. (2021). We restrict the sample to developing countries as in our main regression analysis.

Based on a panel regression within country over time, Wacziarg and Welch (2008) find that the liberalization events raise openness, defined similarly as in our study, by 5.53 percentage points of GDP. Table C1 focuses on the $\beta^{DiD}$ impacts on $ETR_K$ and $ETR_L$. Despite being based on broader criteria, the trade liberalization events produce qualitatively similar results to the main event-study (Section 5.1), with positive impacts on both $ETR$s, and a larger magnitude-impact on $ETR_K$ than $ETR_L$. Figure C1 estimates the dynamic event-study. Liberalized and control countries are on parallel trends until the event onset; both $ETR$s start to increase in the immediate post-event years. In Panel B, the results hold when the control group is formed within a development group (using the full set of categories by the World Bank). Panel C shows the results are robust to excluding countries which have cross-border capital liberalization events at any point during the sample-period (Bekaert, Harvey and Lundblad, 2000). Finally, Panel D shows the results hold when we exclude countries with concurrent domestic reforms (Wacziarg & Wallack, 2004).

Figure C1: Event-Study of Trade Liberalization Based on Wacziarg & Welch (2008)

(a) Effective tax rate on capital (b) Effective tax rate on labor




*Notes:* These graphs show event-study impacts of the trade liberalization events from Wacziarg and Welch (2008) on $ETR_K$ (left panel) and $ETR_L$ (right panel).

Table C1: Trade Liberalization Event-Study Based on Wacziarg & Welch (2008)

| | $ETR_K$ | $ETR_L$ |
|---|---|---|
| | (1) | (2) |
| Panel A: *Benchmark* | | |
| Imputed treatment effect | 0.062*** | 0.025*** |
| | (0.022) | (0.007) |
| $N$ | 3046 | 3046 |
| Panel B: *With development category-year fixed effects* | | |
| Imputed treatment effect | 0.062*** | 0.025*** |
| | (0.022) | (0.007) |
| $N$ | 3046 | 3046 |
| Panel C: *Excluding countries with capital liberalization* | | |
| Imputed treatment effect | 0.047** | 0.015* |
| | (0.023) | (0.008) |
| $N$ | 2314 | 2314 |
| Panel D: *Excluding countries with domestic reforms* | | |
| Imputed treatment effect | 0.063*** | 0.018** |
| | ((0.022) | (0.007) |
| $N$ | 2782 | 2782 |

*Notes:* This table shows the results from estimating the difference-in-difference regression and the imputed treatment effect of the 68 trade liberalization events from Wacziarg and Welch (2008), between 1965 and 2008. The sample is developing countries (as defined in Section 3.3). The imputed treatment effect is based on the method in Borusyak, Jaravel, and Spiess (2021). In Panel B, the estimation is augmented with interactive fixed effects between year and development category; the category distinguishes between upper middle income countries, and lower middle income and low income countries. In Panel C, the sample excludes all countries that have a capital liberalization reform at any point during the sample-period, based on Bekaert, Harvey and Lundblad (2000). In Panel D, the sample excludes all countries with domestic reforms which coincide in timing with their trade liberalization event, based on Wacziarg and Wallack (2004). Standard errors are clustered at the country level. For more details on the liberalization events, see Appendix C.3. * p<0.10 ** p<0.05 *** p<0.01.

# Appendix D Results on Tax Capacity Mechanism

## D.1 Relating our results to the trade and public finance literatures

In Table 1, a one hundred percentage point increase in trade openness causes $ETR_K$ to rise by 14.9 percentage points (column 2). To what extent is this estimate comparable with previous work in international trade and public finance on trade's impact on firm size and on the firm size gradient in effective tax rates? Based on equation (6), the tax capacity hypothesis argues that trade can impact $ETR_K$ through two main channels:

- (A) Formalization channel: Trade can change the country's corporate sector share:

$$\frac{d\ ETR_K}{d\ \mu^K_{Corp}} \frac{d\ \mu^K_{Corp}}{d\ trade} = \frac{d\ \mu^K_{Corp}}{d\ trade} * (ETR^K_{Corp} - ETR^K_{NC}) \tag{11}$$

- (B) Firm size channel: Trade can impact firm size ($y$) within the corporate sector:

$$\frac{d\ ETR_K}{d\ y} \frac{d\ y}{d\ trade} \approx \mu^K_{Corp} * \frac{d\ ETR^K_{corp}}{d\ y} \frac{d\ y}{d\ trade} \tag{12}$$

In what follows, we obtain effect sizes of channels (A) and (B) based on the trade and public finance literatures. Our main calibration shows that, for a 100 percentage points increase in trade to GDP, the effect on $ETR_K$ via channel (A) is 9.8 pp and via channel (B) is 1.5 pp. Together, these two effects imply a 11.3 pp increase in $ETR_K$, which accounts for 75% of the 14.9pp coefficient estimated in Table 1.

**Channel A** This channel considers how trade causes an increase in the corporate share of the economy. In turn, since the average effective tax rate is higher in the corporate sector than in the non-corporate sector, the increase in the corporate share leads to an increase in the overall effective tax rate $ETR_K$.

We draw on Dix-Carneiro et al. (2024). The authors conduct a counterfactual exercise which shows how the size of the formal sector and firm size respond to changes in international trade costs. The authors do not directly model the corporate and non-corporate sector; they instead model the share of workers in formal and informal firms, which we use as our proxy. Moreover, the authors do not directly show the results of trade costs on firm size; they show results on TFP, which we will use as a proxy for firm size. The estimation results of interest are in Table 8. In this counterfactual, the authors analyze the responses in the model to a reduction in iceberg trade costs (from 2.4, the benchmark, to 2.3). The authors do not directly report the magnitude of the change in trade openness that results from the 0.1 unit change. However, in the following sub-section, the authors discuss that a reduction in trade costs from 2.4 to 2.0 causes an increase in the imports-GDP ratio from 6.6% to 15%. Assuming linearity, this would imply that a 0.1 unit decrease in trade costs causes a 2.1 percentage point increase in imports/GDP. Abstracting from any changes in exports, this translates into a 2.1 percentage point increase in trade openness.

In Table 8, the authors report results separately for manufacturing (the 'C' sector) and services (the 'S' sector). Let us assume that each sector has an equal share in the economy.

All estimates are expressed as percent changes relative to the benchmark with trade costs that equal 2.4. The average percent change in informal employment is -2.875 (with a 4.4 percent reduction in manufacturing and a 1.31 percent reduction in services). To obtain a percentage point change, we refer to the right hand panel of Figure 3, which shows that the informal employment share is 51% at a level of trade cost of 2.4. In turn, the 0.1 unit decrease in trade costs leads to a 1.5 percentage points reduction in the informal share – or, equivalently in our setting, a 1.5 percentage points increase in the corporate share.

Taken together, these numbers imply that a 100 percentage point increase in trade openness leads to a 71.4 percentage point increase in the corporate share.

In Section 6, we report that $ETR^K_{Corp}$ is 19.8, which is 46 percent larger than the overall $ETR^K$ at 13.6. In turn, using the corporate share of national income in 1990 as our benchmark (Figure 1), which is 55%, this implies that $ETR^K_{NCorp}$ is 6.02. We use the corporate share of national income as a proxy for the corporate share of capital income.

We can plug these numbers into equation (11), which leads to a 9.84 percentage point increase in $ETR^K$: $(19.8 - 6.02) \times 0.714 = 9.84$.

Apart from Dix-Carneiro et al. (2024), there are other relevant papers in the literature, including Coar, Guner, and Tybout (2016), Bernard, Jensen, and Kortum (2003), Eaton, Kortum, and Kramarz (2011), Aw, Roberts, and Xu (2011) and Bernard et al. (2007). Using similar calculations, for a 100 percentage point increase in trade, the model counterfactuals in these papers predict an increase in average firm size between 59 percent and 90 percent.

**Channel B** $ETR^K_{corp}$ can be decomposed into a statutory tax rate, $\tau$, a legally defined tax base $\pi$ (typically economic profits from which some tax incentives can be deducted), and a compliance rate $c$ – the share of the tax base that is reported. This gives $ETR^K_{corp} = \tau \times \pi \times c$. We can define the legal effective tax rate as $\tau \times \pi$ (which, in the absence of tax evasion, is the effective tax rate). The impact of trade on $ETR^K_{corp}$ has two components:

$$\frac{d\ ETR_K}{d\ ln(y)} \frac{d\ ln(y)}{d\ trade} \approx \mu^K_{Corp} \times [\tau \times \pi \times \frac{d\ c}{d\ ln(y)} \frac{d\ ln(y)}{d\ trade} + c \times \frac{d\ \tau.\pi}{d\ ln(y)} \frac{d\ ln(y)}{d\ trade}] \tag{13}$$

For the impact of trade on firm size, the estimation results of interest are in Table 8 and the accompanying section of Dix-Carneiro et al. (2024). The average percent change in TFP is 1.56 percent (with a 1.63 percent increase in manufacturing and a 1.50 percent increase in services). We assume that the percent changes in TFP equates to the percent changes in firm size. These numbers imply that a 100 percentage point increase in trade openness leads to a 74.5 percent rise in firm size.

For the impact of firm size on the compliance rate, we rely on Bachas et al. (2019), BFJ2019 hereafter. The main estimation of BFJ2019 shows that a 10 percentile increase in the size-rank of an industry leads to a 2.2 pp increase in the tax compliance rate (Table 3, panel B, Col 5). An average over the coefficients of developing countries –which corresponds to low and middle income countries in our paper–implies that a 10 percentile increase in the size-rank of an industry leads to a 3.2 pp increase in compliance (Table 4). The average firm-size of an industry doubles from an industry ranked at the 25th percentile to the 50th percentile, and doubles again from the 50th percentile to the 75th percentile (See Table 1, Panel C, Columns 4,5,6). If we convert these numbers around the median size, a 25 percentile rise in the size-rank corresponds to a doubling in firm size, and would lead to a $2.5 \times 3.2 = 8$ pp rise in tax compliance in the corporate sector. To assess how the firm size compliance channel contributes to $ETR_K$, we need to multiply the 8pp compliance rate increase by the legal tax rate on corporate income (35%) and the share of value added in the corporate sector (55% average over the period in LMICs): $8 \times 35\% \times 55\% = 1.54pp$ increase in $ETR_K$ for a doubling of firm size. Based on Dix-Carneiro et al. (2024), a 100 pp increase in trade leads to around a 75% increase in firm size within the corporate sector. This implies that the firm-size compliance channel can explain about a $1.54 \times 75\% = 1.15pp$ increase in $ETR_K$.

Finally, for the impact of firm size on the legal effective tax rate, we rely on Bachas et al. (2025). The authors measure the size gradient in legal effective tax rates for 16 developing countries based on corporate income tax returns. The legal effective tax rate is the tax rate after accounting for all the features of the tax code that permit the tax rate to be lower than the top statutory corporate income tax rate. It accounts for tax provisions, such as reduced rates for smaller firms and for specific sectors, tax incentives and tax exemptions.

The paper directly maps the legal tax rate as a function of firm size in multiple developing countries (Figure 2). This shows that in the majority of countries the $ETR$ increases with firm size within the formal corporate sector, except at the very top of the distribution, and is mainly due to reduced rates provided for firms below a given size. Table A5 of the paper provides a regression coefficient that moves across percentiles of firm size: the average is a coefficient of 0.05. This means that gaining 20 percentile ranks in firm size leads to a 1pp increase in the legal effective corporate tax rate ($ETR_{K,Corp,legal}$). At the median firm size, a 20 percentile rank increase corresponds to about a doubling of firm size. In practice, to determine the contribution of this channel to $ETR_K$, we need to multiply this number by the compliance rate of these firms (which we take to be 80% based on the World Bank Enterprise Surveys used in BFJ2019) and the share of the corporate sector's value-added (55%): this means that a doubling of firm size can account for a $1 \times 80\% \times 55\% = 0.45pp$ increase in $ETR_K$ via the higher legal effective tax rate. In practice, a 100 pp increase in trade leads to around a 75% increase in firm size within the corporate sector (Dix-Carneiro et al., 2024). This implies that the firm-size legal effective tax rate channel causes a 0.35 pp increase in $ETR_K$.

Taken together, in channel (B) the effect of trade on $ETR_K$ via an increase in firm size in the corporate sector causes about a 1.5 pp increase in the $ETR_K$.

An alternative, relevant paper is Best et al., 2021(version used), which studies evasion rates, measured via random audits, as a function of firm size in Pakistan. The paper shows that tax compliance rates are strongly correlated with firm size. The results from their paper imply that doubling firm size is associated with a 4.4-12.2 pp increase in compliance.

## D.2 Firm-level analysis in Rwanda

**Data-sources and sample** Our analysis draws on three administrative datasets from the Rwanda Revenue Authority (RRA), for the years 2015-2017. These data sources can be linked through unique tax identifiers for each firm, assigned by the RRA for the purpose of collecting customs, corporate income and value-added taxes. The first data source is the customs records, which contain information on international trade transactions made in each year by each firm. We use this data to measure each firm's direct imports. The second data is the firms' corporate income tax (CIT) declarations merged with the firm registry. These data contain detailed annual information on firms' profits, revenue and costs. We use these data to measure each firm's effective tax rate. The third data source is the business-to-business transactions database. These data are retrieved through the electronic billing machines (EBM) that all firms registered for VAT are legally required to use (Eissa and Zeitlin, 2014). For a given seller, EBMs record the transactions to each buyer identified by the tax firm-ID. We use this data to measure buyer-seller relationships.

When combined, these data allow us to construct the buyer-supplier relationships of the Rwandan formal economy and document firms' total trade exposure. Importantly, since the network data is based on tax-IDs, we cannot observe transaction linkages with informal, non-registered firms. This sample selection on formal firms also features in most recent network studies, by virtue of relying on administrative data, including in Chile (Huneeus, 2020); Costa Rica (Alfaro-Ureña et al.); Ecuador (Adao et al., 2022); India (Gadenne et al., 2022); Turkey (Demir et al., 2021); and Uganda (Almunia et al., 2023).

Our sample is the set of firms that are registered for CIT. We do not restrict the sample to having positive capital income in the CIT data; doing so would be consequential, as a significant number of registered CIT firms are 'nil filers' that report zero income or sales. Keen (2012) describes how 'nil filers' are common in developing countries. We measure each firm $i$'s yearly effective tax rate on corporate profits, corresponding to corporate $ETR_i^K$ in equation (6), as the ratio of corporate taxes paid divided by net profit. Net profit is revenue minus material, labor, operational, depreciation and financial costs. In order to not condition the sample on positive net profits, we add 1 to the denominator. We include in the sample the set of firms that pay tax based on reported turnover rather than reported profits.

**Exposure to trade** To measure a firm's total exposure to trade, we follow Dhyne et al. (2021) who use similar administrative datasets as ours to measure trade exposure of Belgian firms. We define firm $i$'s total foreign input share as the share of inputs that it directly imports ($s_{Fi}$), plus the share of inputs that it buys from its domestic suppliers $l$ ($s_{li}$), multiplied by the total import shares of those firms:

$$s_i^{Total} = s_{Fi} + \sum_{l \in V_i} s_{li} \cdot [s_{Fl} + \sum_{r \in V_l} s_{rl} \cdot (s_{Fr} + ...)]$$

where $V_i$ is the set of domestic suppliers of firm $i$, and $V_l$ is the set of domestic suppliers of firm $l$. The denominator of the input shares is the sum of purchases from other firms and imports. Note that $s_i^{Total}$ is recursive: a firm's total foreign input share is the sum of its direct foreign input share and the share of its inputs from other firms, multiplied by those firms' total foreign input shares. We limit the calculation to the inputs from a firm's immediate suppliers $l$ as well as the suppliers to their suppliers $r$ (adding more network-levels only marginally increases $s_i^{Total}$). $s_i^{Total}$ reflects the direct import share of firm $i$'s suppliers and the suppliers' suppliers, each weighted by the share of inputs that each firm buys from other domestic firms. We focus on firms' exposure to imports through their supplier network.

The data reveals that while just under 30% of firms import directly, 93% rely on trade either directly or indirectly through their suppliers. In the median firm, the total foreign input share is 48% (it is 39% for the median Belgian firm in Dhyne et al., 2021).

**Impacts of trade exposure on $ETR^K$ and size** We investigate the association between trade exposure ($s_i^{Total}$) and $ETR_i^K$ in a regression form in Table 4. The outcome differs across columns: in columns (1)-(2) and (5)-(6), it is the effective tax rate on corporate profits, $ETR_i^K$; in columns (3)-(4), it is annual revenue. All regressions are estimated based on the Poisson pseudo-maximum likelihood estimator, with standard errors clustered at the firm level. In columns (1) to (4), trade exposure $S^{Total}$ is instrumented with trade-shocks to firms and their supplier network based on the shift-share design of Hummels et al. (2014). As Poisson pseudo maximum likelihood does not currently, to the best of our knowledge, allow for an IV estimation, we estimate 2SLS and report bootstrapped standard errors. The $1^{st}$-stage F-statistic is reported in the middle part of the table. In columns (5)-(6), revenue is directly regressed on $ETR_i^K$. Given the estimation method, revenue is standardized in columns (5)-(6) to ease comparison with the other columns. All columns include firm fixed effects. Odd-numbered columns include year fixed effects, while even-numbered columns include year×industry fixed effects, year×geography fixed effects and firm-year controls (the firm's total number of suppliers and clients).

The main results are described in Section 6.3. Because the estimation is within the corporate sector, this exercise cannot speak to the magnitude of trade's net impact on sector-level $\overline{ETR}_C^K$. These firm-level results on corporate $ETR_i^K$ are therefore complementary to the country-level results on $\overline{ETR}_C^K$. An additional limitation is that the network linkage measures are derived from administrative data which, by construction, only exist for tax registered firms (Atkin & Khandelwal, 2020). This sample restriction implies that this firm-level regression is not suited to study the impacts of trade on the size of informal firms.

## D.3 Type of trade analysis

We investigate whether trade has differential impacts on $ETR$ and mechanism outcomes depending on the nature of the trade variation (Section 6.4). We use our two instruments to investigate the impacts of: (i) imports versus exports (of trade in both intermediate G-S and final G-S); (ii) trade in intermediate G-S versus final G-S (summed across imports and exports). We use UN's Broad Economic Categories (Rev. 5) to classify final versus intermediate goods-services (G-S), combining capital goods with the latter. For the imports versus exports IV, the two $1^{st}$-stage regressions are

$$log(imp_{ct}) = \beta_1 \cdot Z_{ct}^{gravity} + \beta_2 \cdot Z_{ct}^{oil-dist} + \mu_c + \mu_t + \epsilon_{ct}$$
$$log(exp_{ct}) = \pi_1 \cdot Z_{ct}^{gravity} + \pi_2 \cdot Z_{ct}^{oil-dist} + \eta_c + \eta_t + \iota_{ct}$$

where $log(imp_{ct})$ and $log(exp_{ct})$ are the logs of total imports to NDP and total exports to NDP, respectively, in country $c$ in year $t$. The log-transformation improves the $1^{st}$-stage, as does the use of NDP weights. The $2^{nd}$-stage is

$$y_{ct} = \theta_1 \cdot log(imp_{ct}) + \theta_2 \cdot log(exp_{ct}) + \kappa_c + \kappa_t + \phi_{ct}$$

The set-up is similar for the second IV (intermediate G-S vs final G-S) where we replace $log(imp_{ct})$ and $log(exp_{ct})$ with the log of total trade in intermediate G-S to NDP and the log of total trade in final G-S to NDP. IV results for developing countries are in Panel A of Table D1, with $1^{st}$-stage regressions in Panel B. Note that it is ex ante unclear if the two instruments generate a strong overall first-stage. We gauge this by inspecting the Kleibergen-Paap F-statistics, which are not well above conventional threshold levels (13.57 and 3.69). Given this challenge, we limit our scope to studying whether the coefficient signs for the different types of trade are consistent with our simplified predictions (and whether they statistically differ from each other). The exclusion restriction requires that the regressors add up to total trade openness. For this reason, we cannot implement an IV which focuses on the impacts of final versus intermediate G-S for, say, imports only. This also implies that, for a given outcome, the hypotheses in our two IVs (final versus intermediate G-S; imports versus exports) will be correlated. We accordingly adjust the p-values for multiple hypotheses testing using the Romano-Wolf method.

The results are described in Section 6.4. Since we only have 2 instruments, we cannot decisively conclude on the impacts for the 4 types of trade (imports of intermediate G-S, exports of intermediate G-S, imports of final G-S, exports of final G-S). Notwithstanding, the estimated IV coefficients are consistent with imports of final G-S decreasing $ETR_K$ and mechanism outcomes ($\mu_C$,$\overline{ETR}_C^K$), and imports of intermediate G-S increasing them.

## Table D1: Type of Trade Analysis in Developing Countries

| **Panel A: IV** | $ETR_K$ | | $ETR_L$ | | Corporate Income | | Mixed Income | | Corporate $ETR_K$ | |
|---|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
| Export of G-S | 0.487* | | 0.225** | | 0.214 | | -0.159 | | 0.612* | |
| | (0.263) | | (0.097) | | (0.135) | | (0.096) | | (0.339) | |
| | [0.053] | | [0.039] | | [0.072] | | [0.125] | | [0.066] | |
| Import of G-S | -0.359*** | | -0.184*** | | -0.127* | | 0.069 | | -0.442*** | |
| | (0.127) | | (0.045) | | (0.074) | | (0.049) | | (0.159) | |
| | [0.053] | | [0.039] | | [0.072] | | [0.158] | | [0.066] | |
| Intermediate G-S | | 0.304*** | | 0.133*** | | 0.147** | | -0.119** | | 0.385*** |
| | | (0.095) | | (0.039) | | (0.071) | | (0.046) | | (0.122) |
| | | [0.046] | | [0.039] | | [0.026] | | [0.046] | | [0.046] |
| Final G-S | | -0.245*** | | -0.125*** | | -0.089** | | 0.050** | | -0.303*** |
| | | (0.051) | | (0.023) | | (0.045) | | (0.025) | | (0.057) |
| | | [0.019] | | [0.019] | | [0.006] | | [0.099] | | [0.006] |
| F-test: Equality of coefficients [p-value] | 4.82 | 14.78 | 8.56 | 19.06 | 2.73 | 4.34 | 2.55 | 5.98 | 4.60 | 15.36 |
| | [0.030] | [0.001] | [0.004] | [0.000] | [0.101] | [0.039] | [0.113] | [0.016] | [0.034] | [0.000] |
| $N$ | 3750 | 3803 | 3750 | 3803 | 3750 | 3803 | 3750 | 3803 | 4572 | 4572 |

| **Panel B: $1^{st}$-stage** | Import of G-S | Export of G-S | Intermediate G-S | Final G-S |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| $Z^{gravity}$ | 0.287*** | 0.253*** | 0.282*** | 0.269*** |
| | (0.035) | (0.060) | (0.035) | (0.053) |
| $Z^{oil-distance}$ | 0.077*** | 0.002 | 0.008 | 0.116*** |
| | (0.011) | (0.018) | (0.014) | (0.020) |
| $1^{st}$-stage F-statistic | 134.51 | 15.76 | 54.77 | 75.85 |
| $1^{st}$-stage Sanderson-Windmeijer Weak Instrument F-statistic | 36.50 | 34.02 | 65.35 | 70.61 |
| $1^{st}$-stage Kleibergen-Papp F statistic | 3.69 | | 13.57 | |
| $N$ | 3750 | 3750 | 3803 | 3803 |

*Notes:* The sample is developing countries ( defined in Section 3.3). Panel A presents IV results, while Panel B presents $1^{st}$-stage results. In Panel A's odd-numbered columns, imports and exports are the regressors while in even-numbered columns it is trade in intermediate goods and services (G-S) and trade in final G-S. Outcomes differ across columns in Panel A: in cols. (1)-(2), effective tax rate on capital, $ETR_K$; in cols. (3)-(4), effective tax rate on labor, $ETR_L$; in cols. (5)-(6), corporate income share of net domestic product; in cols. (7)-(8), mixed income share of net domestic product; in cols. (9)-(10), average effective tax rate on corporate profits. For details on the outcomes and the instruments, see Table 1 and 3. Relative to those tables, the drop in sample size in this table is due to availability of the type of trade classification. For each coefficient, we report in brackets the p-values which correct for multiple hypotheses testing, using the Romano-Wolf method. Multiple hypothesis testing is accounted for within each outcome between the two IV estimations (exports and imports; final G-S and intermediate G-S). At the bottom of each column in Panel A, we report the F-test for the equality of coefficients. In Panel B, cols. (1)-(2) correspond to the first-stage regression that instruments simultaneously for imports and exports; cols. (3)-(4) is the first-stage regression which instruments simultaneously for intermediate G-S and final G-S. In Panel B, we report the F-statistic of excluded instruments; the Sanderson-Windmeer multivariate F-test of excluded instruments; and, the Kleibergen-Paap F-statistic. * p<0.10 ** p<0.05 *** p<0.01. Standard errors in parentheses are clustered at the country level. For more details, see Section D.3.

# Appendix E Capital Liberalization Events

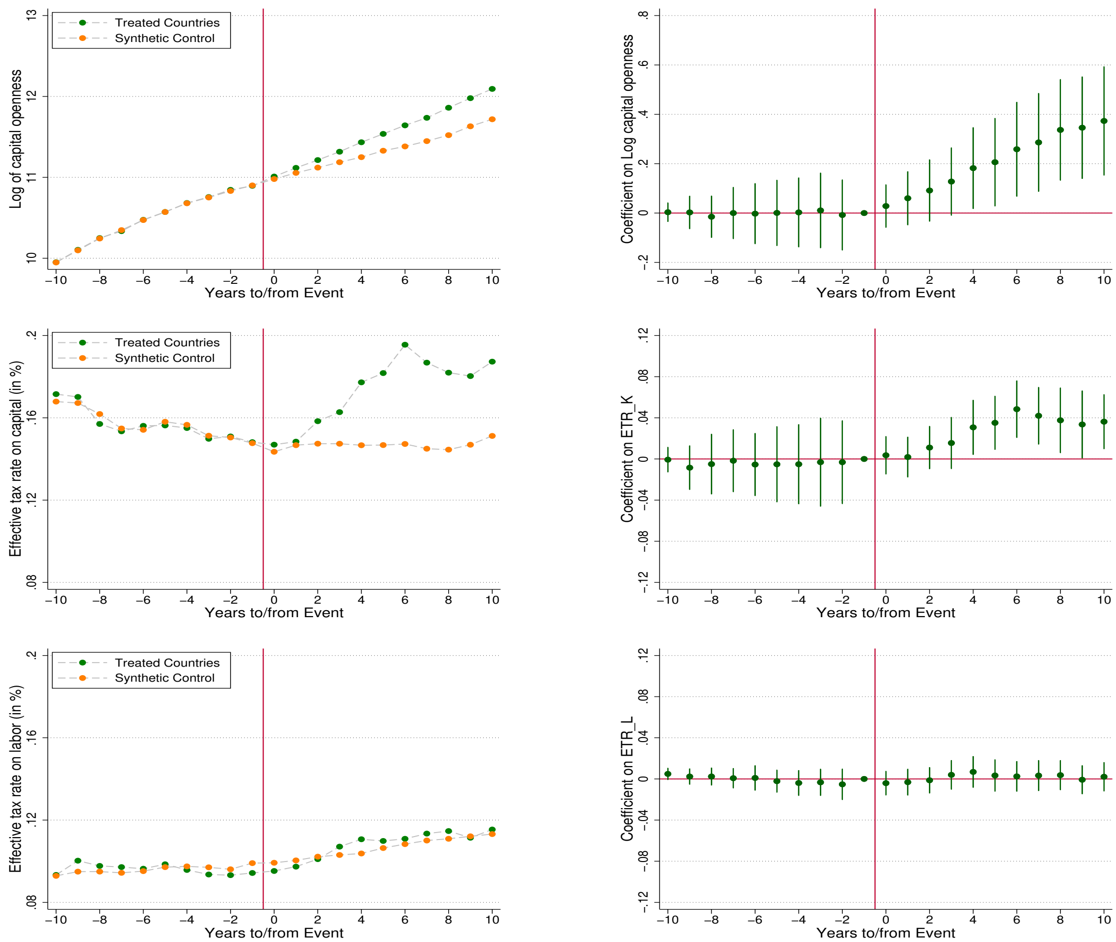
To attempt to investigate the impact of capital liberalization on effective tax rates, we draw on Chari et al. (2012). The authors measure capital liberalization events in 25 developing countries as the date when foreign investment in the domestic stock market was first allowed. They show that these events significantly increase foreign capital inflows, including foreign direct investment (FDI) and import of capital goods.[69] Compared to other policies aimed at lifting FDI restrictions, liberalizing the domestic stock market occurs at a precise point in time, is not marked by policy-reversal or net capital outflow, and is unambiguously related to capital liberalization (Eichengreen, 2001). We employ the empirical design of Section 5.1 and Figure 6. We create a synthetic control country for each of the 25 treated countries and for each outcome. We measure capital openness as the total sum of the stocks of foreign assets and liabilities (Gygli et al., 2019).

Figure E1 reports the event-study results. Relative to a stable pre-trend, we observe a sustained rise in capital openness precisely at the time of the event. $ETR_K$ also increases, with a small lag to the timing of the capital liberalization event; in the medium-run, the positive effect on $ETR_K$ is significant at the 5% level. There is no discernible effect on $ETR_L$. Similar to the reasoning for the trade tax-capacity mechanism, the inflow of foreign capital, as well as any subsequent increase in capital goods imports and aggregate investment, may positively impact $ETR_K$ by contributing to general growth of firms or by causing an expansion of initially larger firms. Consistent with this interpretation, we find that the capital liberalization events led to increases in the corporate output share and the average corporate effective tax rate (results not shown).

One important limitation is that the events considered here remove restrictions on capital *inflows* and are not informative of the impacts of increased capital *outflows*. In general, more work is needed to understand the determinants of policies that impact cross-border capital flows in developing countries and their effects on $ETR$s.

[69]FDI includes green field investments (building plants from scratch) and cross-border mergers and acquisitions (M&A). Chari et al. (2012) note that M&A is impacted by stock market liberalization, makes up to 40-60% of FDI in developing countries, and can trigger subsequent green field investments.

Figure E1: Event Study of Capital Liberalization Reforms




*Notes:* These panels show event-studies for capital liberalization reforms in the 25 developing countries of Chari, Henry, and Sasson (2012). The panels correspond to different outcomes: capital openness (top panels); effective tax rate on capital (middle panels); effective tax rate on labor (bottom panels). Capital openness is the total sum of the stocks of foreign assets and liabilities, in constant USD. We use the log transformation for this outcome. We employ the empirical design of Section 5.1 and Figure 6. The left-hand graphs show the average level of the outcome in every year to/since the event, for treated countries and for synthetic control countries. The right-hand graphs show the estimated $\beta_e$ coefficients on the to/since dummies, based on equation (4) but where the trade liberalization events are replaced with capital liberalization events. The bars represent the 95% confidence intervals. Standard errors are clustered at the country level. Details are in Appendix E.