Impulse Responses By Pseudo-Panel Local Projections: An R Package

Samson M'boueke[†]

This version: August 17, 2025

Abstract

The estimation of impulse responses to shocks across multiple countries is usually of interest in the international economics literature. However, international datasets are often characterized by limited time-series coverage, which undermines the precision of country-specific impulse responses. Researchers often resort to panel regressions to exploit cross-sectional variation and shrink standard errors, but panel regressions impose extensive homogeneity restrictions on the data. This paper develops an R package that generalizes and extends the "Pseudo-Panel Local Projections" strategy described in Berg, Curtis, and Mark (2023) for shrinking impulse-response standard errors without relying on the standard panel regression techniques. While Berg, Curtis, and Mark (2023)'s method relies on constrained estimation of small sets of pseudo-panels based on similar-sized local projection coefficients within horizons, this paper proposes an extension where pseudo-panels are formed by searching for similarities both within and across horizons. This extension increases the chances of obtaining pseudo-panels with enough units to leverage cross-sectional variation and further improve accuracy. An application to the analysis of the effects of commodity terms-of-trade shocks shows that the extended pseudo-panel local projections are better at shrinking impulse response standard errors than Berg, Curtis, and Mark (2023)'s original strategy.

Keywords: Impulse Responses; Local Projections; Pseudo Panels; R Package

[†]Work in progress. Comments are welcome at samsonmboueke@gmail.com

1 Introduction

Since Sims (1980) popularized the analysis of impulse responses in macroeconomics, economists have extensively used this tool to investigate the dynamic effects of economic shocks on various macroeconomic variables. In international macroeconomics, in particular, researchers are often interested in how the effects of these shocks differ across countries. A natural way of addressing such questions is to perform a country-by-country analysis of impulse responses. One issue with international data, however, is that they often come in low frequency (usually annual) and are typically too short, which undermines the precision of country-specific impulse responses. To address these limitations, researchers have often turned to panel regression techniques to exploit cross-sectional variation and shrink impulse response standard errors. However, these techniques impose extensive homogeneity restrictions across countries which can usually not be justified, and country-specific impulse responses are still of interest.

In a recent working paper that estimates the cross-country distribution of the impulse responses of real GDP per capita growth to global and idiosyncratic temperature shocks, Berg, Curtis, and Mark (2023) describe a strategy to shrink country-specific impulse response standard errors in the context of local projections, without relying on the standard panel regression techniques. They do so by constrained estimation of small sets of pseudo-panels for countries with similar-sized local projection estimates at fixed horizons. The method is referred to as "Pseudo-Panel Local Projections". The objective of this paper is threefold: (1) First, I propose a generalization of the pseudo-panel local projections strategy so that researchers can apply it in different contexts and for different units (countries, regions, firms, etc.). (2) I then propose an extension to further shrink impulse response standard errors. The extension relies on the search for units with similar-sized local projection estimates both within and across horizons. This technique generally increases the pool of similar elements in a pseudo-panel, providing more cross-sectional information to use in order to shrink standard errors. (3) Finally, I develop an R package named PsPLP (Pseudo-Panel Local Projections) which mechanically applies these techniques in local projection settings where researchers can leverage cross-sectional variation to improve statistical power.

The PsPLP package provides an array of functions to estimate impulse responses using (a) the standard single-equation local projections method, (b) the generalized version of Berg, Curtis, and Mark (2023)'s pseudo-panel local projections, and (c) the extended pseudo-panel local projections. An application to the analysis of commodity terms-of-trade shocks shows that the extended pseudo-panel local projections method is better at shrinking impulse response standard errors than (a) and (b).

The rest of the paper is structured as follows: Section 1 presents the econometric framework, Section 2 introduces the PsPLP package and its functionalities, Section 3 presents an application of the methods described in Section 1, and Section 4 provides concluding remarks.

2 Econometric Framework

2.1 Impulse Responses with Single-Equation Local Projections

Given a time series Y_t, the local projection equation estimating the response of Y_t at horizon $h \in \{0, \cdots, H\}$ to shocks in $X_{1,t}, X_{2,t}, \ldots, X_{S,t}$ (where S is the number of shock variables) and controlling for variables $C_{1,t}, C_{2,t}, \ldots, C_{K,t}$ (where K is the number of control variables), can be represented as:

$$Y_{t+h} = \alpha_h + \sum_{s=1}^{S} \beta_{s,h} X_{s,t} + \sum_{k=1}^{K} \left(\sum_{\ell=1}^{L_k} \gamma_{k,\ell,h} C_{k,t-\ell} \right) + \varepsilon_{t+h} \tag{1}$$

where Y_{t+h} is the dependent variable at $t+h$, which can be either the simple lead of a variable of interest (y_{t+h}) or its long-difference ($y_{t+h} - y_{t-1}$) at horizon h. α_h is the intercept specific to horizon h, and $\beta_{s,h}$ represents the impulse response of Y to a shock in $X_{s,t}$ at horizon h under reasonable exogeneity assumptions about $X_{s,t}$. When $Y_{t+h} = y_{t+h}$, $\beta_{s,h}$ is specific to period h. When $Y_{t+h} = y_{t+h} - y_{t-1}$, $\beta_{s,h}$ is the cumulative response from $t-1$ to $t+h$. $C_{k,t-\ell}$ is the ℓ^{th} lag of the k^{th} control variable, and the optimal lag lengths $(L_1, \cdots, L_K)$ are allowed to vary across different control variables. ε_{t+h} is the error term at $t+h$.

The PsPLP package estimates the above equation by OLS. The optimal lag lengths are selected using Akaike's Information Criterion (AIC) or the Bayesian Information Criterion (BIC) at horizon $h = 0$, according to the following procedure:

Lag Length Selection in PsPLP: (1) For each control variable $C_{k,t}$, define a range of potential lag lengths (1 to $L_{\max}$), where $L_{\max}$ is the maximum allowable lag length for any control variable in the model. (2) For each $C_{k,t}$, estimate separate local projection models at $h = 0$ for each potential lag length ℓ within its range. Each model includes all the contemporaneous shock variables $X_{1,t}, X_{2,t}, \ldots, X_{S,t}$ and the lags of $C_{k,t}$ up to ℓ. (3) Calculate AIC/BIC and select the lag length L_k^* that minimizes AIC/BIC.

The package also offers the option to correct for potential autocorrelation and/or heteroskedasticity in the error terms with Newey and West (1987) standard errors.

2.2 The Pseudo-Panel Local Projections Method

This paper is the first attempt at generalizing the small-scale pseudo-panel local projections method introduced by Berg, Curtis, and Mark (2023) to shrink impulse response standard errors while keeping point estimates close to the single-equation local projection estimates. The strategy is originally applied in a multi-country local projection setting where the limited availability of time series undermines the precision of single-equation impulse response estimation, and extensive pooling is unjustified. In this section, I attempt to generalize the strategy, and in the next, I propose an extension to further improve precision. The strategy can be described in two steps: (1) the first step consists in forming small panels for each unit at each horizon by pooling together units with similar-sized impulse response point estimates, (2) and in the second step, the corresponding panel local projections are estimated using appropriate panel regression techniques.

Step 1: *Finding Similar Units and Forming Pseudo-Panels*

Suppose you estimate Equation (1) for N units. At the core of the pseudo-panel local projections approach to forming and estimating pseudo-panels is the assessment of similarity between different units (countries/regions/firms, etc.). This is achieved through the comparison of the impulse response point estimates for each unit at each horizon. Specifically, begin with Unit 1 at horizon h. Then using the single-equation local projection estimates, apply a similarity criterion to determine whether there exist any other units similar to Unit 1. The current version of the PsPLP package uses direct comparison of the impulse response coefficients[1] for the relevant shock after appropriately rounding them (see Appendix B, helper function `round_coeffs`, for details on the rounding method). Next, form Unit 1's pseudo-panel[2] consisting of Unit 1 itself and the desired number of units similar to Unit 1. The current version of PsPLP includes all the similar units determined through the selected similarity criterion and keeps only the current unit when there are no similarities. Finally, repeat the process for units 2 through N.

[1]In the original method by Berg, Curtis, and Mark (2023), the authors have two "shock variables" in their local projection equations, an idiosyncratic temperature I_t and a global temperature G_t. Their similarity criterion is as follows: compute the sum of squared errors (SSE) for a country i relative to the remaining countries, defined as $\left(\beta^G_{i,h} - \beta^G_{j,h}\right)^2 + \left(\beta^I_{i,h} - \beta^I_{j,h}\right)^2, j \neq 1$, where $\beta^G_{i,h}$ $(\beta^I_{i,h})$ is the impulse response at horizon h to a shock in G_t (I_t). Then sort by SSE from lowest to highest and select between two to five countries with the lowest SSEs.

[2]Berg, Curtis, and Mark (2023) use the term "pseudo-panel" rather than panel because the number of units in each panel can change from one horizon to the next.

Step 2: *Estimating Pseudo-Panel Local Projections*

After forming the pseudo-panels in Step 1, the corresponding panel local projections will be estimated using appropriate panel regression techniques. Indexing the members of a given pseudo-panel by j, a typical pseudo-panel local projection is of the form:

$$Y_{j,t+h} = \alpha_h + \sum_{s=1}^{S} \beta_{s,h} X_{j,s,t} + \sum_{k=1}^{K} \left(\sum_{l=1}^{L_k} \gamma_{k,l,h} C_{j,k,t-\ell} \right) + \varepsilon_{j,t+h} \tag{2}$$

In the current version of PsPLP, all coefficients are constrained to be common across all units in a given pseudo-panel[3], and Equation (2) is estimated by Generalized Method of Moments (GMM) for pseudo-panels with more than one unit, and by OLS for single-unit pseudo-panels.

2.3 Extended Pseudo-Panel Local Projections

In the pseudo-panel local projections method, instances in which no similarities are found for a specific unit at a specific horizon are not unlikely. When such cases do occur, there is no cross-sectional variation to use, hence no shrinkage. Forming pseudo-panels anyway, for example by selecting the closest units in terms of similarity, will move the new impulse response point estimate too far from the single-equation point estimate, which may further increase standard errors.

The idea of the "extended" pseudo-panel local projections method is that we can increase the chances of obtaining pseudo-panels with at least 2 elements to leverage cross-sectional variation. This can be achieved by searching for similarities not only within horizons but also across horizons. For example, if the impulse response of Unit 1 at horizon h_1 is similar to that of Unit 2 at horizon h_2, then the (Unit 2 - Horizon h_2) combination will be a candidate for Unit 1's pseudo-panel at horizon h_1. This means

[3] In the original method by Berg, Curtis, and Mark (2023), only the coefficients corresponding to the shocks are constrained to be equal.

that in Unit 1's pseudo-panel at horizon h_1, the dependent variable for Unit 2 will be Y_{t+h_2}, and that of Unit 1 will be Y_{t+h_1}. The formal procedure is described as follows:

Step 1: *Finding Similar Unit-Horizon Combinations and Forming Pseudo-Panels*

Begin with Unit 1 at horizon h. Then using the single-equation local projection estimates, apply a similarity criterion across units and horizons to determine whether there exist any other unit-horizon combinations similar to Unit 1-Horizon h. Next, form Unit 1's pseudo-panel consisting of Unit 1-Horizon h itself and the desired number of unit-horizon combinations similar to Unit 1-Horizon h. The current version of PsPLP includes all the similar unit-horizon combinations determined through the selected similarity criterion, and again, keeps only the current unit-horizon combination when there are no similarities. Finally, repeat the process for units 2 through N.

Step 2: *Estimating "Extended" Pseudo-Panel Local Projections*

After forming the pseudo-panels in Step 1, the corresponding panel local projections will be estimated using appropriate panel regression techniques. Indexing the members of a given pseudo-panel by the (j, h_j) couple, where j indexes the unit and h_j its associated horizon, the "extended" pseudo-panel local projection is of the form:

$$Y_{j,t+h_j} = \alpha_h + \sum_{s=1}^{S} \beta_{s,h} X_{j,s,t} + \sum_{k=1}^{K} \left(\sum_{\ell=1}^{L_k} \gamma_{k,\ell,h} C_{j,k,t-\ell} \right) + \varepsilon_{j,t+h_j} \tag{3}$$

In the current version of PsPLP, all coefficients are constrained to be common across all units in a given pseudo-panel, and Equation (3) is estimated by GMM for pseudo-panels with more than one unit-horizon combination, and by OLS for pseudo-panels with a single combination.

3 The PsPLP Package

The package provides a suite of functions designed to estimate the dynamic effects of exogenous shocks on economic indicators using one of three methods: the single-equation, the pseudo-panel, and the "extended" pseudo-panel local projections. The three main functions corresponding to each of these methods are `estimate_LP`, `estimate_panels1`, and `estimate_panels2`.

The `estimate_LP` function estimates the standard single-equation local projections by OLS and should be specified as follows:

```
results <- estimate_LP(df, id_var, time_var, dv_vars, shocks, ctrls,
              max_h, max_lag, use_NW, is_cum, info_criteria)
```

The user must prepare the data (`df`) in a panel data format, which must include identifiers for units (`id_var`) and time (`time_var`). The package allows for multiple dependent variables, which should be specified as `dv_vars = c("var1", "var2", ...)`. The shocks and control variables corresponding to each dependent variable are passed to a list in the same fashion as follows:

```
shocks = list(var1 = c("var1_shock1", "var1_shock2", ...), var2 =
              c("var2_shock1", "var2_shock2", ...), ...)

ctrls = list(var1 = c("var1_ctrl1", "var1_ctrl2", ...), var2 =
              c("var2_ctrl1", "var2_ctrl2", ...), ...)
```

The user also specifies the maximum allowable lag length for any control variable in the model (`max_lag`) for the optimal lag selection process described in subsection 2.1, as well as the maximum horizon for impulse responses (`max_h`). Finally, the user must specify whether to use Newey and West (1987) standard errors (default is No: `use_NW = FALSE`), whether the dependent variable should be in cumulative form (default is No: `is_cum =`

FALSE),[4,5] and the information criteria for lag length selection (default is `info_criteria = "AIC"`). Appendix B provides details on the helper functions for `estimate_LP`, and Appendix A describes the `plot_IR` function that plots the impulse response functions.

The `estimate_panels1` function estimates pseudo-panel local projections and depends on the results from the `estimate_LP` function. It should be specified as follows:

```
panel_results1 <- estimate_panels1(data, panels, results, id_var, max_h,
                   max_lag, is_cum, use_NW)
```

where `data` is the same `df` as in `estimate_LP`, and `id_var`, `max_h`, `max_lag` are the same as in `estimate_LP` and should take the same values as in `estimate_LP`. Only `use_NW` has a default value here (`FALSE`), but it must still be given the same value as in `estimate_LP` for consistency. The object `results` is the results from the `estimate_LP` function. One required step before using the `estimate_panels1` function is the formation of pseudo-panels, which are gathered in the `panels` object. Appendix A provides details on this step, from the search for similarities with the `check_similarity1` function to the actual pseudo-panel formation with the `form_panels1` function. The function `plot_IR_panel1`, also explained in Appendix A, plots the impulse responses from the pseudo-panel local projections estimation.

The `estimate_panels2` function estimates "extended" pseudo-panel local projections. It depends on the results from the `estimate_LP` function but is independent of `estimate_panels1`. It is estimated in the same way as `estimate_panels1`:

```
panel_results2 <- estimate_panels2(data, panels, results, id_var, max_h,
                   max_lag, is_cum, use_NW)
```

All the inputs are specified in the same way as with the `estimate_panels1` function. The function to check for similarities across units and horizons (`check_similarity2`), the

[4]The control variables may differ depending on whether the dependent variable is in cumulative form or not.

[5]One alternative to `is_cum = TRUE` is to define the dependent variable in cumulative form in the data preparation step before passing it to the `estimate_LP` function, and then set `is_cum = FALSE`.

function to form the pseudo-panels containing unit-horizon combinations (`form_panels2`), and the function to plot impulse responses (`plot_IR_panel2`) are all explained in Appendix A.

4 Application: The spillover effects of "foreign" commodity terms-of-trade shocks

This section applies the pseudo-panel local projection techniques to the econometric setup described in Appendix C. The idea is to estimate the spillover effects of a positive shock to the commodity terms of trade of a country's trading partners on said country's domestic economy. The empirical model[6] is:

$$x_{t+h} - x_{t-1} = \alpha_h + \beta_{0,h}\mu_t^* + \sum_{\ell=1}^{L_\mu} \beta_{\ell,h}\mu_{t-\ell}^* + \sum_{\ell=1}^{L_x} \rho_{\ell,h}\Delta x_{t-\ell} + \epsilon_{t+h} \tag{4}$$

where $\mu_t^* = \Delta tot_t^*$, and tot^* is the country-specific "foreign" commodity terms of trade constructed as the trade-weighted average of trading partners' commodity terms of trade. x_t is alternatively a country's own terms of trade (tot), trade balance-to-GDP ratio (tb), real GDP per capita (y), real private consumption per capita (c), real gross investment per capita (i), or real dollar exchange rate index (rer). All variables are in logs, except the trade balance-to-GDP ratio. The data are annual and span the period from 1980 to 2019 for 88 countries (25 advanced, 44 emerging, and 19 developing). I only consider the 19 developing economies in this application.

[6]The optimal lag lengths are allowed to vary across control variables to use the full functionalities of the PsPLP package.

4.1 Single-Equation Local Projections

I start by estimating the single-equation local projections for each of the 19 developing countries using the `estimate_LP` function. I include all 6 dependent variables. The maximum horizon for impulse responses is 5, the maximum lag for optimal lag length selection is 4, and the lag lengths are selected using BIC.

```
results <- estimate_LP(
  df = data,
  id_var = "Country",
  time_var = "Date",
  dv_vars = c("tot", "tb", "y", "c", "i", "rer"),
  shocks = list(tot = "dtot_star", tb = "dtot_star", y = "dtot_star",
                c = "dtot_star", i = "dtot_star", rer = "dtot_star"),
  ctrls = list(tot = c("dtot_star", "dtot"), tb = c("dtot_star", "dtb"),
               y = c("dtot_star", "dy"), c = c("dtot_star", "dc"),
               i = c("dtot_star", "di"), rer = c("dtot_star", "drer")),
  max_h = 5,
  max_lag = 4,
  use_NW = TRUE,
  is_cum = TRUE,
  info_criteria = "BIC"
)
```

Results for the country Chad (`TD`) and the variable *tot* at Horizon 0 show that the impulse response point estimate is -1.16. The optimal lag length selected for the "foreign" commodity terms-of-trade shock (`dtot_star`) and the growth rate of `TD`'s own commodity terms of trade (`dtot`) is 1.

```
> results$`Unit: TD ; Var: tot ; Horizon: 0`
$Model
Call:
lm(formula = form, data = unit_df)
Coefficients:
 (Intercept)     dtot_star  dtot_star_L1       dtot_L1
   -0.006449     -1.162927      1.588626     -0.163901
$SEs
 (Intercept)    dtot_star dtot_star_L1      dtot_L1
  0.02220277   0.80018364   0.82983793   0.14243000
$R2
[1] 0.1779468
$AdjustedR2
[1] 0.1054127
$SelectedLags
$SelectedLags$dtot_star
[1] 1
$SelectedLags$dtot
[1] 1
```

Next, I use the `plot_IR` function to plot the impulse responses with 95% confidence bands. Figure (1) shows the impulse responses of all dependent variables for Chad. We can see that the confidence bands are pretty large. In the next subsection, I estimate pseudo-panel local projections and compare the results with the ones obtained here.

```
# Plot impulse responses
plot_IR(results, "dtot_star", 0.95)
```

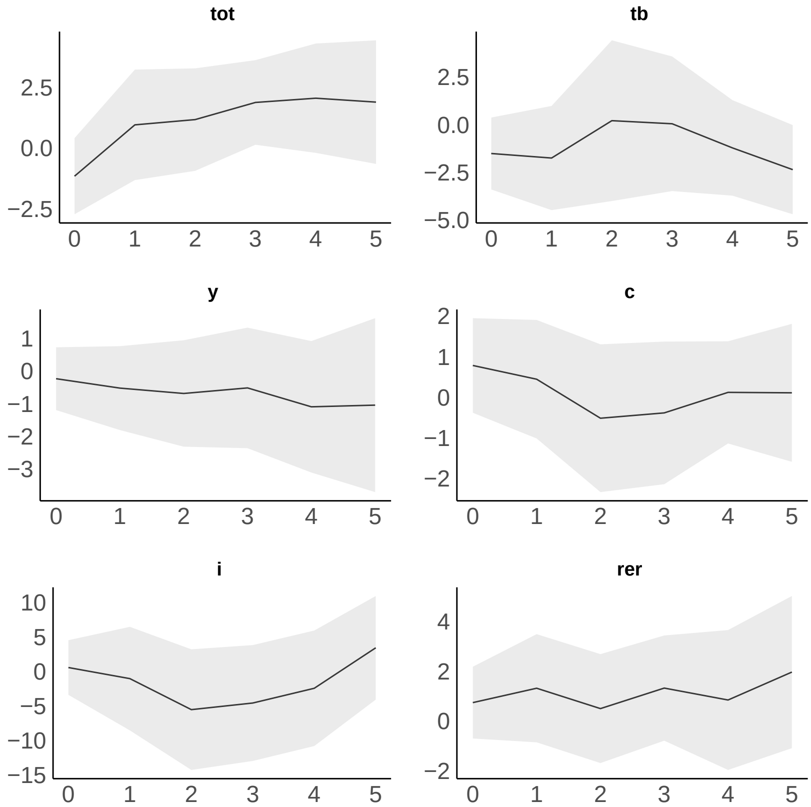

Figure 1: Impulse Responses to a 1 percentage-point increase in the growth rate of the "foreign" commodity terms of trade for Chad: Single-Equation Local Projections.

4.2 Pseudo-Panel Local Projections

I now proceed to the estimation of pseudo-panel local projections. The impulse response point estimates are rounded to 1 significant figure for the similarity check process with `check_similarity1`.

```
# Check for similarities
similarity_results1 <- check_similarity1(results, "dtot_star",
                                          significant_figures =1)
```

Results from the similarity check for Chad, *tot*, and Horizon 2, displayed below, show that 5 countries are similar to Chad in terms of the size of the impulse response point estimate:

```
> similarity_results1$TD$tot$Horizon_2
            Unit_Other Base_Coefficient Other_Coefficient Are_Similar
dtot_star           BD         1.165898        1.07656355         YES
dtot_star11         NE         1.165898        1.17599208         YES
dtot_star13         NP         1.165898        1.41399363         YES
dtot_star15         SN         1.165898        1.25086653         YES
dtot_star16         TG         1.165898        1.45869501         YES
dtot_star1          BF         1.165898        2.41242647          NO
dtot_star2          BI         1.165898       -1.48866858          NO
dtot_star3          CF         1.165898       -0.62778907          NO
dtot_star4          CM         1.165898       -0.47661606          NO
dtot_star5          GH         1.165898       -0.60291085          NO
dtot_star6          GM         1.165898        0.09733354          NO
dtot_star7          GW         1.165898        0.87302423          NO
dtot_star8          KE         1.165898        4.29973778          NO
dtot_star9          MG         1.165898       -0.47398118          NO
dtot_star10         ML         1.165898        2.40433549          NO
dtot_star12         NG         1.165898       -2.00711639          NO
```

Next, I form the panels using the `form_panels1` function. The results for Chad and *tot* at all horizons are displayed below:

```
# Form pseudo-panels
panels1 <- form_panels1(similarity_results1)
```

```
> panels1$TD_tot_0
[1] "TD"
> panel1s$TD_tot_1
[1] "TD"
> panels1$TD_tot_2
[1] "TD" "BD" "NE" "NP" "SN" "TG"
> panels1$TD_tot_3
[1] "TD" "BF" "ML" "NE" "NP" "TZ"
> panels1$TD_tot_4
[1] "TD" "BD" "BF" "ML" "NP"
> panels1$TD_tot_5
[1] "TD" "BF" "ML" "NP"
```

I then estimate the pseudo-panel local projections, plot the resulting impulse response functions for Chad, and compare the results with the single-equation local projections (Figure (2)).

```
panel_results1 <- estimate_panels1(
  data = data,
  panels = panels1,
  results = results,
  id_var = "Country",
  max_lag = 4,
  max_h = 5,
  use_NW = TRUE,
  is_cum = TRUE
)

plot_IR_panel1(panel_results1, "dtot_star", 0.95)
```

The pseudo-panel local projections show some improvement over the single-equation local projections in that the confidence bands look smaller overall. However, in cases where no similarities are found, like in Chad's pseudo-panel for *tot* at horizons 0 and 1, the standard errors do not shrink, and the confidence bands remain the same because the estimated model is the same as in the single-equation local projections. In the next section, I proceed to estimate the "extended" pseudo-panel local projections and compare the results with those of the previous two methods.

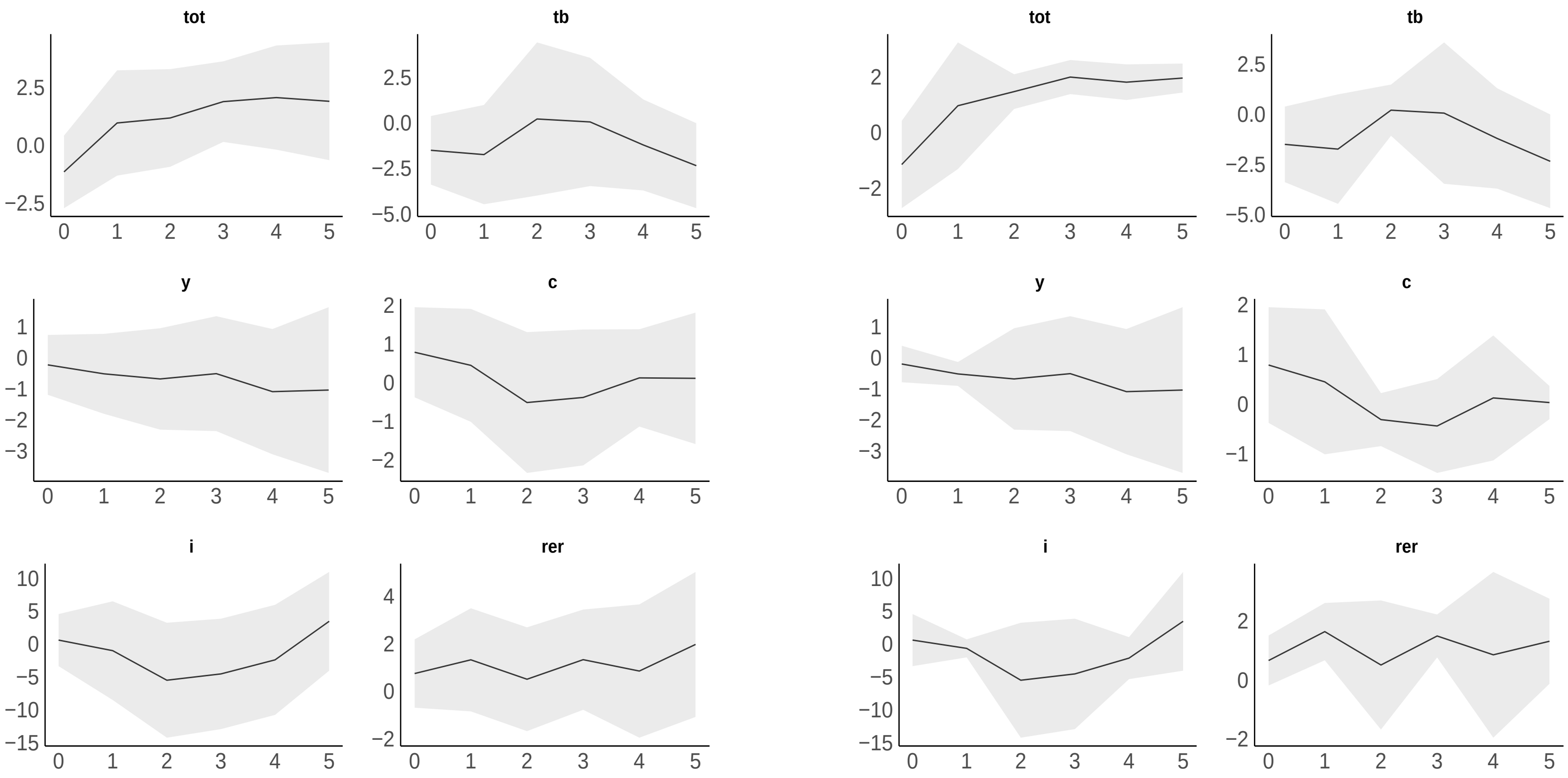

Figure 2: Impulse Responses to a 1 percentage-point increase in the growth rate of the “foreign” commodity terms of trade for Chad: Comparison between Single-Equation (left) and Pseudo-Panel (right) Local Projections.

4.3 "Extended" Pseudo-Panel Local Projections

Finally, I apply the "extended" pseudo-panel local projections method. The results from the pseudo-panel formation step for Chad and *tot* at all horizons are as follows:

```
# Check for similarities
similarity_results2 <- check_similarity2(results, "dtot_star",
                                           significant_figures =1)


# Form pseudo-panels
panels2 <- form_panels2(similarity_results2)

> panel1s2$TD_tot_0
[1] "TD_0" "GH_1" "BI_2" "SD_4" "SD_5"
> panel1s2$TD_tot_1
[1] "TD_1" "GW_2" "GH_3" "SN_3" "TG_3"
> panel1s2$TD_tot_2
[1] "TD_2" "GM_0" "GW_0" "NE_0" "SN_0" "BD_1" "GW_1" "NP_1" "SN_1" "BD_2"
[11] "NE_2" "NP_2" "SN_2" "TG_2" "BD_3" "GH_4" "KE_4" "TZ_4" "TZ_5"
> panel1s2$TD_tot_3
[1] "TD_3" "BD_0" "BF_0" "ML_0" "NP_0" "TG_0" "TG_1" "BF_2" "ML_2" "BF_3"
[11] "ML_3" "NE_3" "NP_3" "TZ_3" "BD_4" "BF_4" "ML_4" "NP_4" "TD_4" "BF_5"
[20] "ML_5" "NP_5" "TD_5"
> panel1s2$TD_tot_4
[1] "TD_4" "BD_0" "BF_0" "ML_0" "NP_0" "TG_0" "TG_1" "BF_2" "ML_2" "BF_3"
[11] "ML_3" "NE_3" "NP_3" "TD_3" "TZ_3" "BD_4" "BF_4" "ML_4" "NP_4"
[20] "BF_5" "ML_5" "NP_5" "TD_5"
> panel1s2$TD_tot_5
[1] "TD_5" "BD_0" "BF_0" "ML_0" "NP_0" "TG_0" "TG_1" "BF_2" "ML_2" "BF_3"
[11] "ML_3" "NE_3" "NP_3" "TD_3" "TZ_3" "BD_4" "BF_4" "ML_4" "NP_4"
[20] "TD_4" "BF_5" "ML_5" "NP_5"
```

We can see that all pseudo-panels for Chad and *tot* have more than one element, and they also have more elements than in the previous approach. The last step is to estimate the pseudo-panels and plot the impulse responses as follows:

```
# Estimate pseudo-panels
panel_results2 <- estimate_panels2(
  data = data,
  panels = panels2,
  results = results,
  id_var = "Country",
  max_lag = 4,
  max_h = 5,
  use_NW = TRUE,
  is_cum = TRUE
)

# Plot impulse responses
plot_IR_panel2(panel_results2, "dtot_star", 0.95)
```

Figure (3) compares the impulse responses from the three methods. Clearly, the "extended" pseudo-panel local projections method yields much smaller standard errors than the two other methods. One caveat, however, is that the "extended" method does not guarantee the formation of pseudo-panels with more than one element. In cases where no similarities are found, there will still be no improvement over the single-equation local projections.

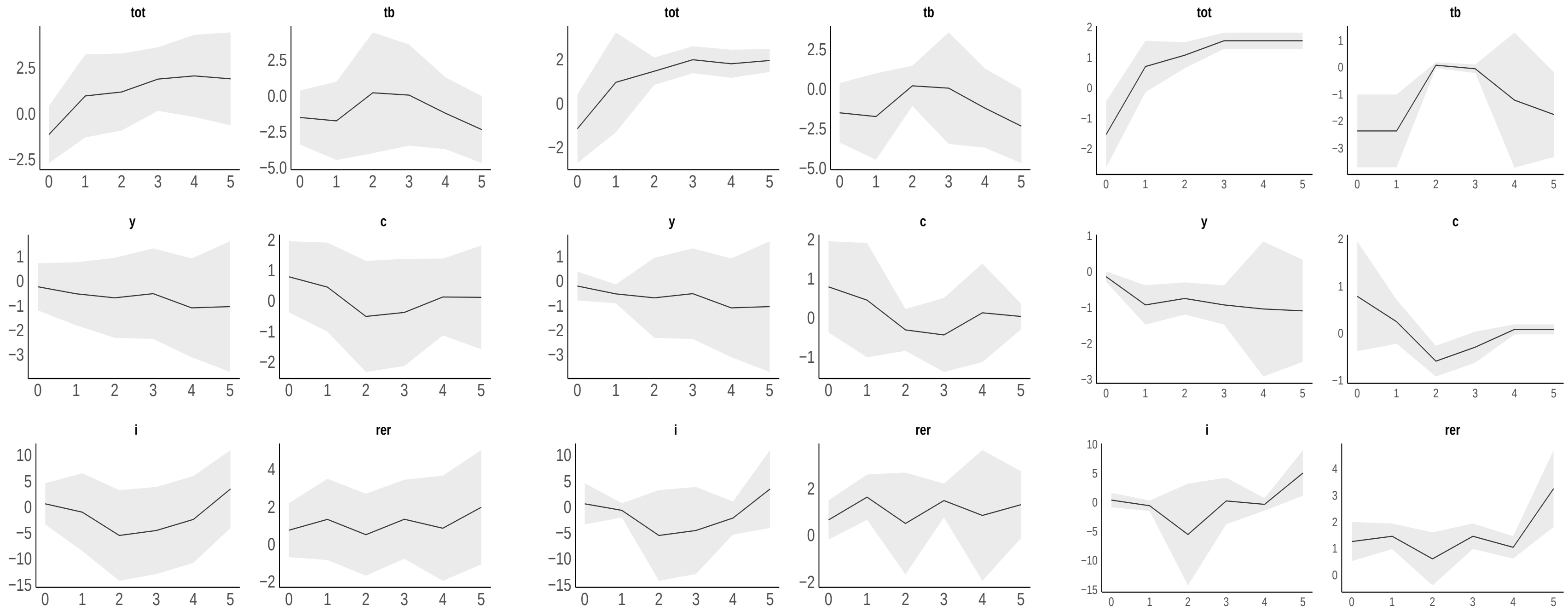

Figure 3: Impulse Responses to a 1 percentage-point increase in the growth rate of the “foreign” commodity terms of trade for Chad: Comparison between Single-Equation (left), Pseudo-Panel (middle), and “Extended” Pseudo-Panel (right) Local Projections.

5 Concluding Remarks

In this paper, I generalize and extend the pseudo-panel local projections method applied in Berg, Curtis, and Mark (2023) to shrink impulse response standard errors while keeping impulse response point estimates close to the single-equation local projection estimates. The original method relies on constrained estimation of small sets of pseudo-panels for units with similar-sized local projection estimates at fixed horizons. The proposed extension is based on the argument that we can further shrink standard errors by increasing the pool of units with similar-sized point estimates in each pseudo-panel, which can be achieved by searching for similarities both within and across horizons. I then develop the PsPLP R package to mechanically apply these techniques in local projection settings where cross-sectional information can be leveraged to improve statistical power. I show via an application that the extended version of the method performs better than both the single-equation local projections and the original pseudo-panel local projections method.

This paper has a few limitations. First, these techniques currently focus on shrinking standard errors but do not look at biases in point estimates. Second, the extended pseudo-panel local projections method does not guarantee the formation of pseudo-panels with more than one element. It just increases the chances of this happening. Third, none of the methods shrinks standard errors in a "uniform" way. The confidence bands might display erratic behavior at times. Finally, the PsPLP is in its basic version with a few restrictions like estimation methods, and is not yet available for download to the public. Future developments will attempt to address these limitations.

References

Berg, Kimberly A, Chadwick C Curtis, and Nelson Mark (2023). *Gdp and temperature: Evidence on cross-country response heterogeneity*. Tech. rep. National Bureau of Economic Research.

Newey, Whitney K. and Kenneth D. West (1987). "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix". In: *Econometrica* 55.3, pp. 703–708. ISSN: 00129682, 14680262. URL: http://www.jstor.org/stable/1913610 (visited on 02/08/2024).

Sims, Christopher A (1980). "Macroeconomics and reality". In: *Econometrica: journal of the Econometric Society*, pp. 1–48.

Appendix

A. Main Functions of the PsPLP Package

`estimate_LP`

Description:

Estimates single-equation local projections to compute unit-specific impulse responses. It handles data preprocessing, model fitting, and the selection of optimal lags based on information criteria.

Usage:

`estimate_LP(df, id_var, time_var, dv_vars, shocks, ctrls, max_h, max_lag, use_NW, is_cum, info_criteria)`

Parameters:

- `df` - Data frame containing the dataset. `df` should be in panel format
- `id_var` - The identifier variable for cross-sectional units in `df`
- `time_var` - The time variable in `df`
- `dv_vars` - The dependent variables of interest in `df`.
- `shocks` - The shock variables of interest in `df`.
- `ctrls` - The control variables in `df`.
- `max_h` - The maximum horizon for impulse responses.
- `max_lag` - The maximum number of lags for any control variable in any unit-specific equation.

- `use_NW` - Boolean flag to use Newey-West standard errors. The default is `use_NW = FALSE`.
- `is_cum` - Boolean flag to use cumulative responses. The default is `is_cum = FALSE`.
- `info_criteria` - Criterion for the selection of optimal lags ('AIC' or 'BIC"). The default is 'AIC'.

`plot_IR`

Description:

Plots and saves impulse response functions from the results obtained by single-equation local projection estimations using the `estimate_LP` function.

Usage:

`plot_IR(results, shock_name, confidence_level)`

Parameters:

- `results` - Results from the `estimate_LP` function.
- `shock_name` - The shock variable of interest. This needs to be one of the shocks used in the `estimate_LP` function
- `confidence_level` - Confidence level for the impulse response functions. The default is 0.95 for 95% confidence bands.

`check_similarity1`

Description:

Checks for similarity of shock effects across different units at fixed horizons. This is done by direct comparison of the impulse response point estimates after rounding to a specified number of significant figures.

Usage:

`check_similarity1(results, shock_name, significant_figures)`

Parameters:

- `results` - Results from the `estimate_LP` function.
- `shock_name` - The shock variable of interest. This needs to be one of the shocks used in the `estimate_LP` function
- `significant_figures` - Number of significant figures to round the point estimates to. The default is 1.

`form_panels1`

Description:

Forms panels based on identified similarities from the `check_similarity1` function. This involves grouping units that showed similar responses to the specified shock.

Usage:

`form_panels1(similarity_results)`

Parameters:

- `similarity_results` - Results from the `check_similarity1` function.

`estimate_panels1`

Description:

Estimates the pseudo-panels formed using the `form_panels1` function. The function uses the GMM method for pseudo-panels with more than one unit and OLS for single-unit pseudo-panels.

Usage:

`estimate_panels1(data, panels, results, id_var, max_lag, max_h,`
`is_cum, use_NW)`

Parameters:

- `data` - The original data frame used for the single-equation local projection estimations. Same as `df` in the `estimate_LP` function
- `panels` - Pseudo-panels formed with the `form_panels1` function.
- `results` - Results from the `estimate_LP` function.
- `id_var` - The identifier variable for cross-sectional units in `data`.
- `max_lag` - The maximum number of lags for any control variable in any unit-specific equation. Logically, the value should be the same as in the single-equation local projection estimation step.
- `max_h` - The maximum horizon for impulse responses. Logically, the value should be the same as in the single-equation local projection estimation step.
- `is_cum` - Whether to calculate cumulative responses. The should be the same as in the single-equation local projection estimation step.
- `use_NW` - Whether to use Newey-West standard errors for OLS estimation. The value should be the same as in the single-equation local projection estimation step for consistency. The default is `use_NW = FALSE`

`plot_IR_panel1`

Description:

Plots and saves impulse response functions from the pseudo-panel local projection results obtained with the `estimate_panels1` function.

Usage:

`plot_IR_panel1(results, shock_name, confidence_level)`

Parameters:

- `results` - Results from `estimate_panels1`.
- `shock_name` - The shock variable of interest. This needs to be one of the shocks used in the estimate LP function
- `confidence_level` - Confidence level for the impulse response functions. The default is 0.95 for 95% confidence bands.

`check_similarity2`

Description:

Checks for similarity of shock effects across units and horizons. This is done by direct comparison of the impulse response point estimates of different unit-horizon combinations after rounding to a specified number of significant figures.

Usage:

`check_similarity2(results, shock_name, significant_figures)`

Parameters:

- `results` - Results from the `estimate_LP` function.
- `shock_name` - The shock variable of interest. This needs to be one of the shocks used in the `estimate_LP` function

- `significant_figures` - Number of significant figures to round the point estimates to. The default is 1.

`form_panels2`

Description:

Forms panels based on identified similarities from the `check_similarity2` function. This involves grouping unit-horizon combinations that showed similar responses to the specified shock.

Usage:

`form_panels2(similarity_results)`

Parameters:

- `similarity_results` - Results from the `check_similarity2` function.

`estimate_panels2`

Description:

Estimates the "extended" pseudo-panels formed using the `form_panels2` function. The function uses the GMM method for pseudo-panels with more than one unit-horizon combination and OLS for pseudo-panels with a single combination.

Usage:

```
estimate_panels2(data, panels, results, id_var, max_lag, max_h,
is_cum, use_NW)
```

Parameters:

- `data` - The original data frame used for the single-equation local projection estimations. Same as `df` in the `estimate_LP` function

- panels - Pseudo-panels formed with the form_panels2 function.
- results - Results from the estimate_LP function.
- id_var - The identifier variable for cross-sectional units in data.
- max_lag - The maximum number of lags for any control variable in any unit-specific equation. Logically, the value should be the same as in the single-equation local projection estimation step.
- max_h - The maximum horizon for impulse responses. Logically, the value should be the same as in the single-equation local projection estimation step.
- is_cum - Whether to calculate cumulative responses. The should be the same as in the single-equation local projection estimation step.
- use_NW - Whether to use Newey-West standard errors for OLS estimation. The value should be the same as in the single-equation local projection estimation step for consistency. The default is use_NW = FALSE

---

plot_IR_panel2

Description:

Plots and saves impulse response functions from the "extended" pseudo-panel local projection results obtained with the estimate_panels2 function.

Usage:

```
plot_IR_panel2(results, shock_name, confidence_level)
```

Parameters:

- results - Results from estimate_panels2.

- `shock_name` - The shock variable of interest. This needs to be one of the shocks used in the estimate LP function

- `confidence_level` - Confidence level for the impulse response functions. The default is 0.95 for 95% confidence bands.

B. Helper Functions in the PsPLP Package

`select_lags`

Description:

Helper function for `estimate_LP`. Determines the optimal number of lags for the shock variables and the other control variables based on the specified information criterion. The lags of the shock variables are included as control variables in the PsPLP package.

Usage:

`select_lags(df, ctrls, max_lag, dv, shocks, info_criteria)`

Parameters:

- `df` - Data frame for a single unit.

- `ctrls` - Control variables for which to select lags.

- `max_lag` - Maximum lag to consider.

- `dv` - A single dependent variable from the list of potential dependent variables to be fed to the `estimate_LP` function.

- `shocks` - Shock variables included in the models.

- `info_criteria` - Information criterion for lag selection.

`gen_lags`

Description:

Helper function for `select_lags`, `estimate_LP`, `estimate_panels1`, and `estimate_panels2`. Generates the lags for the specified variables in the data frame up to the specified maximum lag.

Usage:

`gen_lags(df, vars, max_lag)`

Parameters:

- `df` - Data frame for a single unit.
- `vars` - Variables for which to generate lags.
- `max_lag` - Maximum number of lags to generate.

`gen_leads`

Description:

Helper function for `estimate_panels1` and `estimate_panels2`. Generates leads or long differences for the specified variables.

Usage:

`gen_leads(df, vars, max_h, is_cum)`

Parameters:

- `df` - Data frame for a single unit.
- `vars` - Variables for which to generate leads.
- `max_h` - Maximum horizon for lead variables.

- `is_cum` - Boolean flag to choose between leads and long differences.

`extract_var_names`

Description:

Helper function for `estimate_panels1` and `estimate_panels2`. Extracts the variable names from the models generated in the `results` object from the `estimate_LP` function, ensuring that the pseudo-panel local projections techniques are being applied on the same set of variables as the single-equation local projections.

Usage:

`extract_var_names(results)`

Parameters:

- `results` -Results from the `estimate_LP` function.

`round_coeffs`

Description:

Helper function for `estimate_panels1` and `estimate_panels2`. Rounds a number to a specified number of significant figures.

Usage:

`round_coeffs(number, significant_figures)`

Parameters:

- `number` - Numeric value to be rounded.
- `significant_figures` - Number of significant figures to round to.

If the number is zero, the function returns zero. For non-zero values, it calculates the order of magnitude using the formula:

$$\texttt{order\_of\_magnitude} = \lfloor \log_{10}(\texttt{abs}(number)) \rfloor$$

where $\lfloor \cdot \rfloor$ denotes the floor function, effectively determining the highest power of ten that fits into the absolute value of the number. Next, the number of decimal places needed for rounding is computed as:

$$\texttt{decimal\_places} = \texttt{significant\_figures} - \texttt{order\_of\_magnitude} - 1$$

This step ensures that the final rounded number retains the desired level of significant digits by adjusting the decimal precision accordingly. The number is then rounded to these calculated decimal places.

For example, consider that $number = 0.049$, and we want to round this number to `significant_figures = 1`. The order of magnitude is:

$$\texttt{order\_of\_magnitude} = \lfloor \log_{10}(0.049) \rfloor = -2$$

Hence, the number of decimal places for rounding is:

$$\texttt{decimal\_places} = 1 - (-2) - 1 = 2$$

Therefore, rounding 0.049 to 1 significant figure means rounding it to 2 decimal places, and the result is **0.05**.

C. Econometric Setup for PsPLP Application

I start by constructing the country-specific "foreign" terms of trade indices, borrowing from the global vector autoregressive (GVAR) literature. In the GVAR literature, "foreign" variables are indices calculated from domestic variables so as to match the international trade, financial, or other desired patterns of the country under consideration. This application focuses on foreign variables that match international trade patterns.

The domestic terms of trade variables considered are calculated as the ratio of a commodity export price index to a commodity import price index. For each country i, the "foreign" commodity terms of trade index is calculated, in logs, as the weighted average of the (log) commodity terms of trade of all other countries, using trade weights. That is,

$$tot^*_{i,t} = \sum_{k=1}^{N} w_{i,k} tot_{k,t} \tag{5}$$

where tot is the log of the domestic commodity terms of trade, N is the number of countries, $\sum_{k=1}^{N} w_{i,k} = 1$ and $w_{i,i} = 0$. For each country i, the trade weight $w_{i,k}$ is the share of its trading partner k in its total trade (exports plus imports)[7], calculated by dividing the amount of trade with k by its total trade.[8] It is worth noting that $w_{i,k}$ and $w_{k,i}$ are not the same, because the importance of one country in another country's overall trade depends on the number of trading partners each of them has, and the distributions of trade volumes across partners.

Empirical Model

To obtain the impulse responses of domestic variables to commodity terms-of-trade shocks from trading partners, that is, the spillover effects of "foreign" commodity terms-of-trade shocks, I directly project each country's domestic variables on its corresponding

[7] Total trade here is exports plus imports of all products, not only commodities.

[8] The amount of trade of i with k is the total of the exports from i to k and the imports of i from k, which is different from the amount of trade of k with i.

"foreign" commodity terms-of-trade shocks as follows:

$$x_{t+h} - x_{t-1} = \alpha_h + \beta_{0,h}\mu_t^* + \sum_{\ell=1}^{L_\mu} \beta_{\ell,h}\mu_{t-\ell}^* + \sum_{\ell=1}^{L_x} \rho_{\ell,h}\Delta x_{t-\ell} + \epsilon_{t+h} \tag{6}$$

where x_t is alternatively the domestic commodity terms of trade, the trade balance-to-GDP ratio, real GDP per capita, real private consumption per capita, real gross investment per capita, and the real dollar exchange rate[9]. All variables, except the trade balance-to-GDP ratio, are in logs. μ_t^* is the "foreign" commodity terms-of-trade shock identified as the change in the "foreign" commodity terms of trade from one period to another: $\mu_t^* = \Delta tot_t^* = tot_t^* - tot_{t-1}^*$. The dependent variables are expressed in cumulative terms, implying that the coefficient $\beta_{0,h}$ measures the cumulative impact of the shock.

Data Sources

The trade data used to construct trade weights come from the World Integrated Trade Solution (WITS). For each trading pair, total import and export values are collected using the Harmonized System (HS) 1988/1992 nomenclature. Countries have different data spans, the maximum coverage being 1988-2019. The full dataset contains 177 reporters and 179 partners. For each trading pair, the trade weight is an average of year-specific trade weights.

The raw data used to calculate the dependent variables come from the World Bank's World Development Indicators (September 2023 vintage) and are described in the table below. The study sample consists of countries with at least 30 consecutive observations between 1980 and 2019 for all variables. There are 88 such countries, categorized in Advanced Economies (AEs, 25 countries), Emerging Market Economies (EMEs, 44 countries), and Developing Economies (DEs, 19 countries) according to the International

[9] Before log-transformation, the real exchange rate is calculated as $RER_t = \frac{\mathcal{E}_t P_t^{US}}{P_t}$ where $\mathcal{E}_t$ denotes the dollar nominal exchange rate, given by the domestic-currency price of one U.S. dollar, P_t^{US} denotes the U.S. consumer price index, and P_t denotes the domestic consumer price index

Monetary Fund (IMF)'s classification.

Variables from the World Development Indicators

Variable	Description	Code
NBTT	Net barter terms of trade index (2000=100)	TT.PRI.MRCH.XD.WD
IM	Imports of goods and services (% of GDP)	NE. IMP. GNFS. ZS
EX	Exports of goods and services (% of GDP)	NE. EXP. GNFS. ZS
Y	GDP per capita in constant LCU	NY.GDP. PCAP. KN
C	HHs & NPISHs final cons. exp. (% of GDP)	NE.CON.PRVT.ZS
I	Gross capital formation (% of GDP)	NE. GDI. TOTL. ZS
NER	Official exchange rate (LCU/US$, period avg.)	NY.GDP. PCAP. KN
P	Consumer Price Index (2010 = 100)	FP.CPI.TOTL.

The data covers the period from 1980 to 2019. The trade balance is calculated as $TB = EX - IM$. All variables (except the trade balance) are log-quadratically detrended. The trade balance takes negative values, so it cannot be log-transformed. Instead, it is first divided by the quadratic trend component of output (Y) and the resulting ration is quadratically detrended.