Replication package for "The effect of input price discrimination on retail prices: Theory and evidence from France"

Marie-Laure Allain Claire Chambolle Stéphane Turolla

November 2025

This document provides details on the steps needed to replicate the analysis in the paper "The effect of input price discrimination on retail prices: Theory and evidence from France" by Marie-Laure Allain, Claire Chambolle, and Stéphane Turolla. It also contains a description of the data and programs used.

1 Overview

The replication package is structured into 4 main directories: **Ado**, **Analysis**, **Data_Build**, and **Data_Raw**. To replicate the results, a preamble is to create the 4 folders and save the STATA programs and data as presented in Tables 1 and 2. More precisely:

1. The folder named **Ado** contains the ado-files of user-written programs with the versions needed to replicate the build and analysis files.

2. The folder named **Data_Raw** contains the data needed to replicate the analysis. Except for one database, all data are restricted-access (see details on the data availability in the section below). The folder contains 4 subfolders: **Transactions**, **Products**, **Households**, and **Other**. **Transactions** contains the Kantar Worldpanel (KWP) transaction databases; one file by year. **Products** contains product characteristics databases; one file by year. **Households** contains household characteristics databases ; one file by year. **Other** contains additional data used for the analysis.

3. The folder named **Data_Build** contains three subfolders: **Program**, **Temp**, and **Use**. **Program** contains all the relevant code files needed to create the main analysis datasets. The latter two folders store the relevant intermediate files (in **Temp**) and analysis files (in **Use**).

4. The folder named **Analysis** contains two subfolders: **Program** and **Output**. **Program** contains all the relevant code files needed to replicate the figures and tables in the paper and uses as inputs the files that are produced and stored in **Data_Build/Temp** and **Data_Build/Use**. **Output** contains two subfolders that store the analysis output. All figures are saved in **Output/Figure**, and all tables are saved in **Output/Table**.

Table 1: Structure of the Replication Package (1/2)

Folder	Subfolder	File
Ado	plus	Stores all the .ado files in their corresponding alphabetic subsubfolder
Analysis	Output/Table	will contain all the tables produced by running the code
Analysis	Output/Figure	will contain all the figures produced by running the code
Analysis	Program	descriptive_statistics_by_subcategory_final_sample.do
Analysis	Program	descriptive_statistics_expenditure_by_household_and_month_KWP_and_LME_products.do
Analysis	Program	descriptive_statistics_expenditure_by_individual_KWP_and_LME_products.do
Analysis	Program	descriptive_statistics_market_share_by_brand_type_KWP.do
Analysis	Program	descriptive_statistics_market_shares_by_retail_group_and _by_month_by_year_for_KWP_and_sample.do
Analysis	Program	descriptive_statistics_on_the_final_sample.do
Analysis	Program	descriptive_statistics_on_the_final_sample_expenditure.do
Analysis	Program	descriptive_statistics_share_of_selected_products.do
Analysis	Program	master_analysis.do
Analysis	Program	regression_DiD_definition1.do
Analysis	Program	regression_DiD_definition1_by_subcat.do
Analysis	Program	regression_DID_definition1_definition5_similar_subcat.do
Analysis	Program	regression_DiD_definition1_price_gap.do
Analysis	Program	regression_DiD_definition1_retail_group.do
Analysis	Program	regression_DiD_robustness_alternative_time_frames.do
Analysis	Program	regression_DiD_robustness_comparison_group_definition.do
Analysis	Program	regression_DiD_robustness_paralell_trends_tests.do
Analysis	Program	regression_DiD_robustness_significance_level_f_test.do
Analysis	Program	regression_DiD_robustness_trimming_procedure.do
Analysis	Program	regression_event_study_definition1.do
Analysis	Program	regression_price_variation_by_year_and_NB_PL_KWP.do
Analysis	Program	regression_retailer_NB_price_index.do
Analysis	Program	regression_retailer_PL_price_index.do
Analysis	Program	regression_robustness_great_recession.do
Analysis	Program	regression_robustness_parallel_trend_for_top8_subcategories.do

Table 2: Structure of the Replication Package (2/2)

Folder	Subfolder	File
Data_Build	Program	brand_type.do
Data_Build	Program	correction_old_code_product.do
Data_Build	Program	cpi_deflator.do
Data_Build	Program	data_preparation.do
Data_Build	Program	definition1.do
Data_Build	Program	definition2.do
Data_Build	Program	definition3.do
Data_Build	Program	definition4.do
Data_Build	Program	definition5.do
Data_Build	Program	family_definition.do
Data_Build	Program	household_characteristics.do
Data_Build	Program	label_family.do
Data_Build	Program	label_subcategory.do
Data_Build	Program	master_build.do
Data_Build	Program	reshape_product_characteristics.do
Data_Build	Program	subcategory_definition.do
Data_Build	Temp	will contain all the temporary files
Data_Build	Use	will contain all the databases used for the analysis
Data_Raw	Households	menages_2006V3_nf_rev.dta
Data_Raw	Households	menages_2007V3_nf_rev.dta
Data_Raw	Households	menages_2008_nf_rev.dta
Data_Raw	Households	menages_2009_nf_rev.dta
Data_Raw	Households	label_clas.do
Data_Raw	Households	label_cspc.do
Data_Raw	Households	label_cspp.do
Data_Raw	Households	label_etup.do
Data_Raw	Households	label_nf.do
Data_Raw	Households	label_reve.do
Data_Raw	Households	label_revp.do
Data_Raw	Households	label_trap.do
Data_Raw	Other	CPI_INSEE_2006_2009.csv
Data_Raw	Products	produits_complets_2006.dta
Data_Raw	Products	produits_complets_2007.dta
Data_Raw	Products	produits_complets_2008.dta
Data_Raw	Products	produits_complets_2009.dta
Data_Raw	Transactions	achats_complets_depts_marques_mdd_2006_V2.1.1_STATA12.dta
Data_Raw	Transactions	achats_complets_depts_marques_mdd_2007_V3.1.1_STATA12.dta
Data_Raw	Transactions	achats_complets_depts_marques_mdd_2008_V2.1.1_STATA12.dta
Data_Raw	Transactions	achats_complets_depts_marques_mdd_2009_V1.1.1_STATA12.dta

2 Data Availability and Provenance Statements

This project primarily uses Kantar homescan panel data (i.e., KWP) whose access is restricted for confidentiality and privacy reasons.

2.1 Statement about Rights

- We certify that the authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

2.2 Summary of Availability

The paper relies on restricted-access homescan panel data as well as publicly available data. Below, we provide information on how academic researchers can gain access to the KWP data. The publicly available data correspond to consumer price indices provided by the French National Institute of Statistics (INSEE) and are included in our replication package.

2.3 Details on each Data Sources

- **Kantar Worldpanel**: This paper uses Kantar homescan panel data (Kantar Worldpanel, 2009) for the years 2006-2009. The authors accessed these data through a partnership between INRAE, the home research institute of two of them, and Kantar. KWP data are made available to INRAE agents (subject to conditions) via the Odalim INRAE platform but remain the property of Kantar Worldpanel. INRAE staff members wishing to use the KWP data for their research must submit a request to `odalim@inrae.fr`. For academic researchers outside INRAE, the data are available for purchase from Kantar. Updated contact information for Kantar France can be found here: `https://www.kantar.com/fr`.

  The KWP data contain information on daily food purchases for home consumption, including online purchases, by household enrolled in the panel. They consist of three databases:

  - Transaction database: each observation corresponds to the purchase of an item by a household in a given store at a specific time. Each record contains information on the components of the purchase (household, product, store, expenditure, etc.).
  - Product database: each observation corresponds to a product, described by several characteristics.
  - Household database: each observation corresponds to a household, described by several characteristics.

- **Supplementary Data**: The paper uses publicly available data on the Consumer Price Index (CPI) for France for the period 2006-2009. We rely on the CPI provided by INSEE to deflate transaction prices. The CPI is calculated for all French households and covers the prices of all products and services. The base year is 2015, and the series identifier is 001759970. These data are included in the **Data_Raw/Other** folder of the replication package. The series was downloaded from `https://www.insee.fr/fr/statistiques/serie/001759970#Telechargement`

3 Data List

3.1 Raw Data

Table 3: List of Raw Data Files

Data Name	Data File (in Data_Raw/)	Source	Provided
KWP	Transactions/achats_complets_depts_marques_mdd_2006_V2.1.1_STATA12.dta	KWP	No
	Transactions/achats_complets_depts_marques_mdd_2007_V3.1.1_STATA12.dta	KWP	No
	Transactions/achats_complets_depts_marques_mdd_2008_V2.1.1_STATA12.dta	KWP	No
	Transactions/achats_complets_depts_marques_mdd_2009_V1.1.1_STATA12.dta	KWP	No
	Products/produits_complets_2006.dta	KWP	No
	Products/produits_complets_2007.dta	KWP	No
	Products/produits_complets_2008.dta	KWP	No
	Products/produits_complets_2009.dta	KWP	No
	Households/menages_2006V3_nf_rev.dta	KWP	No
	Households/menages_2007V3_nf_rev.dta	KWP	No
	Households/menages_2008_nf_rev.dta	KWP	No
	Households/menages_2009_nf_rev.dta	KWP	No
Deflators	Other/CPI_INSEE_2006_2009.csv	INSEE (2025)	Yes

3.2 Analysis Data

Table 4: List of Analysis Data Files

Data File (in Data_Build/Use/)	Usage	Note
base_purchases_2006_2009_Definition_1_F_Test_1pct.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_F_Test_1pct _Robustness_No_Trimming.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_F_Test_1pct _Robustness_Trimming_2SD.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_F_Test_1pct _Robustness_Trimming_309SD.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_Lincom_Trend_1pct.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_Lincom_Trend_1pct _Expenditure.dta	DiD regressions	Transaction
base_purchases_2006_2009_Definition_1_Lincom_Trend_5pct.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_Lincom_Trend_5pct _Expenditure.dta	DiD regressions	Transaction
base_purchases_2006_2009_Definition_1_F_Test_2pct.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_F_Test_3pct.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_F_Test_4pct.dta	DiD regressions	Chain-product-month
base_purchases_2006_2009_Definition_1_F_Test_5pct.dta	DiD regressions	Chain-product-month
Did_estimates_subcategory.dta	DiD regressions	Subcategory
Test_parallel_trends_hypothesis_pre_merger_Definition_1_F_Test_1pct _Robustness_Trimming_BoxPlot.dta	DiD regressions	Chain-product-month
weight_product_in_total_expenditure.dta	DiD regressions	Product
weight_subcat_in_total_expenditure.dta	DiD regressions	Subcategory

4 Computation Requirements

4.1 Software Requirements

Stata code was run using Stata MP/18.0. The code uses the following Stata user-written programs that are stored in the folder **Ado**:

- `distinct.ado` (as of 2010-01-07)
- `egenmore.ado` (as of 2025-09-03)
- `estadd.ado` (as of 2025-09-03)
- `estfe.ado` (as of 2025-09-03)
- `estout.ado` (as of 2025-09-03)
- `estpost.ado` (as of 2025-09-03)
- `eststo.ado` (as of 2025-09-03)
- `esttab.ado` (as of 2025-09-03)
- `fcollapse.ado` (as of 2025-09-02)
- `fegen.ado` (as of 2025-09-02)
- `fegen_group.ado` (as of 2025-09-02)
- `fisid.ado` (as of 2025-09-02)
- `flevelsof.ado` (as of 2025-09-02)
- `fmerge.ado` (as of 2025-09-02)
- `fsort.ado` (as of 2025-09-02)
- `ftab.ado` (as of 2025-09-02)
- `ftools.ado` (as of 2025-09-02)
- `fjoin.ado` (as of 2025-09-02)
- `local_inlist.ado` (as of 2025-09-02)
- `labmask.ado` (as of 2018-11-18)
- `ms_add_comma.ado` (as of 2025-09-02)
- `ms_compile_mata.ado` (as of 2025-09-02)
- `ms_expand_varlist.ado` (as of 2025-09-02)
- `ms_fvstrip.ado` (as of 2025-09-02)

- `ms_fvunab.ado` (as of 2025-09-02)
- `ms_getversion.ado` (as of 2025-09-02)
- `ms_parse_absvars.ado` (as of 2025-09-02)
- `ms_parse_varlist.ado` (as of 2025-09-02)
- `ms_parse_vce.ado` (as of 2025-09-02)
- `nsplit.ado` (as of 2016-06-03)
- `parallel_map.ado` (as of 2025-09-02)
- `parallel_map_template.do.ado` (as of 2025-09-02)
- `reghdfe+.ado` (as of 2025-09-02)

4.2 Memory and Runtime Requirements

The codes were run on a 2021 90 Go PC AMD EPYC Processor 3.19 GHz running Windows 10. The approximate time needed to fully replicate all codes and reproduce the analysis on this machine is 19 hours.

4.3 Description of Program/Code

The file `master_build.do` in the folder **Data_Build/Program** uses the raw data and prepares all the intermediary data needed for the analysis.

The file `master_analysis.do` in the folder **Analysis/Program** uses the output produced by `master_build.do` and replicates all figures and tables in the paper and Online Appendix.

4.4 Instructions to Replicators

4.4.1 Constructing Datasets

To produce all intermediate and cleaned data files used in the analysis, it is sufficient to run `master_build.do`. The code first uses a shell command to clear and re-create the **Data_Build/Temp** and **Data_Build/Use** folders. To run this do-file, the restricted-access KWP data must be added to the raw folders, as described above. Once all data are saved in their appropriate folder, replicators need just to change the directory path before running this do-file:

1. The global specifying the path on the initial code lines in `master_build.do`

All other paths are stated relative to the main directory and should remain unchanged. In `master_build.do`, the do-files are run in the necessary order. Replicators should also note that some do-files are called upon within other do-files, so not all the files in **Data_Build/Program** in the replication package will be listed in `master_build.do`.

4.4.2 Analyzing Data

To produce the paper's figures and tables, replicators must first ensure that all analysis files have been created. Then, execute all analysis codes by running `master_analysis.do` in the **Analysis/Program** folder after adjusting the directory path. The code first uses a shell command to clear and re-create the **Analysis/Output** folders. It then runs all analysis codes in the correct order to produce the results in the paper.

4.4.3 List of Tables/Figures and Programs

Table 5: List of Tables Produced by Program

Table #	Program (Analysis/Program/)	Output file (Analysis/Output/Table)
2	regression_price_variation_by_year_and_NB_PL_KWP.do	Table_2_before_and_after.tex
3	descriptive_statistics_on_the_final_sample.do	Table_3_descriptive_statistics_final_sample.txt
4	regression_DiD_definition1.do	Table_4_regression_DID.tex
4.1 (Section 6.1, in the text)	regression_DID_definition1_definition5_similar_subcat.do	Table_4_1_complementary_analysis_NB_PL_common_subcat.tex
4.1 (Section 6.1, in the text)	regression_DID_definition1_definition5_similar_subcat.do	Table_4_1_complementary_analysis_NB_PLD_common_subcat.tex
5	descriptive_statistics_expenditure _by_household_and_month_KWP_and_LME_products.do	Table_5_expenditure_by_household_brand_type_preperiod _income_decile_KWP_and_LME_products.txt
5.1 (Section 6.2, in the text)	descriptive_statistics_expenditure _by_household_and_month_KWP_and_LME_products.do	Table_5_1_expenditure_by_household_brand _type_pre_period_KWP_and_LME_products.txt
6	regression_DiD_definition1_price_gap.do	Table_6_regression_DiD_split_by_price_gap.tex
7	regression_DiD_definition1_retail_group.do	Table_7_regression_DiD_split_by_retail_group.tex
A1	descriptive_statistics_market_shares_by_retail_group	Table_A1_Market_Shares_by_Retail_Group_and_by_Month_by_Year_for_Final_Sample.txt
A1	_and_by_month_by_year_for_KWP_and_sample.do	Table_A1_Market_Shares_by_Retail_Group_and_by_Month_by_Year_for_KWP.txt
A2	descriptive_statistics_by_subcategory_final_sample.do	Table_A2_descriptive_statistics_by_subcat.txt
A3	regression_DiD_robustness_alternative_time_frames.do	Table_A3_regression_DiD_robustness_alternative_time_frame.tex
A4	regression_DiD_robustness_trimming_procedure.do	Table_A4_regression_DiD_robustness_trimming_procedure.tex
A5	regression_DiD_robustness_significance_level_f_test.do	Table_A5_regression_DiD_robustness_significance_level_f_test.tex
A6	regression_DiD_robustness_paralell_trends_tests.do	Table_A6_regression_DiD_robustness_paralell_trends_selection.tex
A7	regression_DiD_robustness_comparison_group_definition.do	Table_A7_regression_DiD_robustness_alternative_comparison_groups.tex
A8	regression_robustness_great_recession.do	Table_A8_robustness_great_recession.tex

Table 6: List of Figures Produced by Program

Table #	Program (Analysis/Program/)	Output file (Analysis/Output/Figure)
1	regression_event_study_definition1.do	Figure_1_event_study.png
2	regression_DiD_definition1_by_subcat.do	Figure_2_heterogeneous_price_effect_subcat.png
3	regression_retailer_NB_price_index.do	Figure_3_retailer_NB_price_index_weighted.png
A1(a)	regression_robustness_parallel_trend_for_top8_subcategories.do	Graph_monthly_trend_coef_definition1_subcategory_29.png
A1(b)		Graph_monthly_trend_coef_definition1_subcategory_86.png
A1(c)		Graph_monthly_trend_coef_definition1_subcategory_145.png
A1(d)		Graph_monthly_trend_coef_definition1_subcategory_300.png
A1(e)		Graph_monthly_trend_coef_definition1_subcategory_372.png
A1(f)		Graph_monthly_trend_coef_definition1_subcategory_400.png
A1(g)		Graph_monthly_trend_coef_definition1_subcategory_405.png
A1(h)		Graph_monthly_trend_coef_definition1_subcategory_406.png

References

INSEE (2025). "Consumer Price Index". Version (Accessed July 9, 2025). In: URL: https://www.insee.fr/fr/statistiques/serie/001759970#Telechargement.

Kantar Worldpanel (2009). *Extraction via Hindex*.