The Market for Inflation Risk: Supplemental online appendix

Saleem Bahaj
University College London
& Bank of England

Robert Czech
Bank of England

Sitong Ding
London School of Economics

Ricardo Reis
London School of Economics

April 2025

1 UK inflation and the swap market indices

Section 1.1 explains how we adjust the data for the timing of the calculation of the cash flows for inflation swaps. Section 1.2 compares RPI and CPI inflation for the UK.

1.1 Adjusting swap prices for indexation lags

In this section of the appendix, we provide more details for how we adjust the UK RPI swap prices for indexation lags to extract the component that is purely forward-looking.

There are two relevant market conventions for RPI swap pricing. In the first, the floating rate index on which liabilities of the floating payer are calculated is referenced to the RPI index from two months ago. This is known as the RPI indexation lag. Second, UK RPI inflation swap pricing uses a *monthly* RPI fixing. This implies that regardless of which date the swap is traded at a given month, its reference RPI index is always the RPI index from two months ago.

To see the impact of these, consider a 1-year zero coupon RPI swap. If it is traded in a given date in June 2023, this swap has a reference fixing of the April 2023 RPI index, and the swap seller is liable for the floating rate payment that arises from the year-on-year increase of the April RPI index between 2023 and 2024. Because the May (and possibly June) 2023 RPI would have been released at the trade date, the inflation swap seller would require the breakeven rate to compensate for the growth in the RPI index between April and May. This is the known component of the breakeven price. The forward-looking component of this price would then be the expected change in the RPI index between April 2024 and June 2023, reflecting market's expectation of RPI inflation 10-months from now.

Formally, let $p^{(N)}_{t_d,t_m}$ denote the annualized breakeven rate of a N-year zero coupon RPI swap that is traded in date t_d of month t_m. The two time indexes are required because inflation swaps trade at a daily frequency and so their prices vary day-to-day with changes in expected inflation, but the reference RPI index that is tied to the floating leg changes only at a monthly frequency. The breakeven rate at the trade date of t_d in month t_m must satisfy:

$$(1 + p^{(N)}_{t_d,t_m})^N = \frac{RPI^e_{\cdot,t_m+12N-2}}{RPI_{\cdot,t_m-2}} = \underbrace{\left(\frac{RPI^e_{\cdot,t_m+12N-2}}{RPI_{\cdot,t_m}}\right)}_{\textit{unknown payoff}} \underbrace{\left(\frac{RPI_{\cdot,t_m}}{RPI_{\cdot,t_m-2}}\right)}_{\textit{known payoff}}, \tag{1}$$

where $RPI^e_{\cdot,t_m+12N-2}$ denotes the date t_d expectation (in month t_m) of the future RPI realization in month $t_m + 12N - 2$, conditional on the information of the economy at date t_d. Using this equation, we extract the unknown component of the breakeven rate using the daily swap rates and the monthly RPI data from the UK's Office of National Statistics.

The differences in frequencies of the data poses a further set of complications. The latest RPI statistic is released in a lumpy manner, at approximately the middle of each month. As a result, for some swap prices in a given month t_m, the information for RPI of month t_m is not yet available at the trade date. For example, if the RPI is released only on the 16^{th}, swaps executed on dates prior to 16^{th} must price in an unknown component that is attributed to the stochasticity of the RPI in month t_m.

We adjust for indexation by assuming that for all swaps traded in a given month t_m, the month t_m release of the latest RPI figure is not yet available. Hence, our method extracts the known component of the payoff arising only from the change in RPI from month $t_m - 2$ to $t_m - 1$, and regards the indexation-adjusted N-year zero swap price as an annualized rate reflecting RPI inflation expectations over the horizon from month t_m (inclusive) to month $t_m + 12N - 2$.

Therefore, we adjust the formula above to calculate instead:

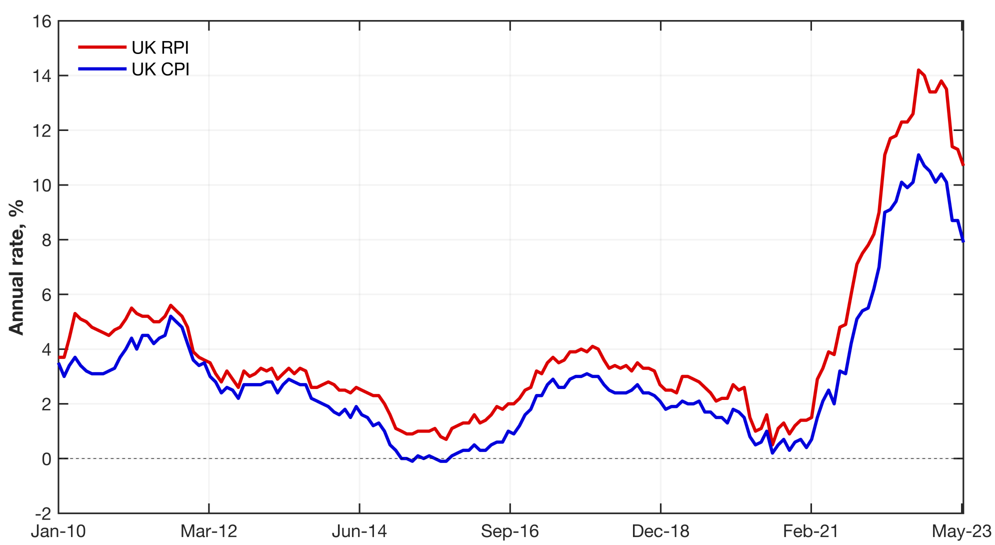
$$(1 + p^{(N)}_{t_d,t_m})^N = \left(\frac{RPI^e_{\cdot,t_m+12N-2}}{RPI_{\cdot,t_m-1}}\right)\left(\frac{RPI_{\cdot,t_m-1}}{RPI_{\cdot,t_m-2}}\right) = \left[(1 + x^{(N)}_{t_d,t_m})^N\right]^{\frac{12N-1}{12N}}\left(\frac{RPI_{\cdot,t_m-1}}{RPI_{\cdot,t_m-2}}\right). \tag{2}$$

The second equality comes from defining $x^{(N)}_{t_d,t_m}$ as the indexation-adjusted swap rate of the N-year zero coupon swap in annualized terms. Taking logs and rearranging, we we solve for the indexation-adjusted N-year zero coupon swap breakeven rate $x^{(N)}_{t_d,t_m}$, for $N = 1, ..., 10, 12, 15, 20, 25, 30, 35, 40, 45, 50$ according to the following equation:

$$\log(1 + x^{(N)}_{t_d,t_m}) = \frac{12N}{12N-1}\left[\log(1 + p^{(N)}_{t_d,t_m}) - \frac{1}{N}\log\left(\frac{RPI_{\cdot,t_m-1}}{RPI_{\cdot,t_m-2}}\right)\right]. \tag{3}$$

1.2 The RPI-CPI wedge

Figure 1 COMPARISON BETWEEN UK RPI AND UK CPI



NOTE: Time-series comparison of the UK RPI and UK CPI. SOURCE: Office for National Statistics.

Inflation swaps are indexed to the Retail Price Index (RPI) while the Bank of England's inflation target refers to the Consumer Price index (CPI). Linking swaps and bonds to RPI is a legacy of earlier periods when RPI was the primary UK price index. Since long-dated bonds from that time are still in circulation, RPI has persisted as the benchmark index in inflation-linked financial markets. In 2020, the UK chancellor announced that RPI is to be aligned with CPIH "no earlier than Feb 2030", with no compensation for holders of index-linked gilts. However, the market is yet to price this transition in its entirety, likely due to expectations of a delay or possible compensation. Given the slow-moving nature of the transition, our estimates of high-frequency movements in RPI swap prices should not be affected by this future change.

Figure 1 plots the two to highlight that they often, but not always move together, and that the RPI is on average higher.

2 Trade repository data and the data cleaning procedures

This section of the Appendix provides a more elaborate documentation of trade repository data used in our analyses, along with a step-by-step procedure taken to carefully construct a data cleaning infrastructure that is capable of handling large volumes of this data from the DTCC trade repository. We then describe the steps taken to collapse the raw high-frequency data into a daily time series that is used across all of our identification strategies. Our goal is to provide some guidance to researchers who are interested in using this novel dataset, as well as to discuss caveats and nuances in the data that all require careful discretion.

2.1 Data structures and nomenclatures: the raw data

Under the UK-EMIR regulation, there are currently 4 trade repositories regulated by the FCA for which derivatives transacted by any UK-based counterparty can be reported to. These are DTCC Derivatives Repository Plc, UnaVista Limited, REGIS-TR UK Limited and ICE Trade Vault Europe Limited. These trade repositories record daily derivative transactions on 5 different asset classes: (i) interest rate (IR), (ii) foreign exchange (FX), (iii) credit (CR), (iv) equity (EQ) and (v) commodity (CO). We focus only on the DTCC trade repository because it is the largest — it captures around 75% to 80% of the market and is therefore sufficiently representative of the data. Within each trade repository, there are two primary types of raw data reports generated in the form of delimited files (e.g., a .txt or .csv file type). These are the (i) trade *state* reports and (ii) trade *activity* reports. Each of these is generated on a daily frequency (with the exception of Sunday, and public holidays) by around noon of each day, to capture the latest derivative transactions from the preceding business day. Each derivative transaction is recorded by a row of contract-level information, spanning more than a 100 columns of variables (The DTCC trade state reports generated after 31st October 2017 have exactly 186 variables). These variables contain information such as the legal entity identifier (LEI) of the reporting counterparty (and its counterparty), trade ID, notional principal (and currency), a series of date fields (such as execution date, effective date, maturity date, settlement date, and termination date), fixed and floating rates, venue of execution, and so forth. See UK EMIR validation rules for a more thorough listing and description of the variables (link).

2.2 Data structures and nomenclatures: trade state reports and trade activity reports

Trade state reports reflect the *stock* of outstanding transactions in the market that had been executed and which have not yet matured. These transactions may have an effective date that precedes the execution date (backward-starting) or one that is scheduled to be at some date in the far future (forward-starting). A daily sequence of trade state reports would therefore grow in the number of transactions, insofar as most of the new derivatives traded have an initial time-to-maturity that is relatively long-dated. Hence, trade state reports are ideal if a researcher would like to obtain empirical facts about how a market has grown over time, or to get a snapshot of the outstanding transactions at a particular instance in history. Trade activity reports capture the *flow* of trading activity from every business day. Therefore, they record the derivative transactions that would be in principle equivalent to the change in the stock of transactions reflected by the trade state files. Thus, the trade activity reports can be ideal for implementing event-based studies, or to assemble a time series data that reflects daily trading in the market. It is, however, also possible to achieve this using trade state reports with an additional caveat that more elaborate deduplication procedures would be required. For instance, an inflation swap contract executed on date T with an initial time-to-maturity of two periods would be present in the trade state reports on *both* dates T and $T+1$ owing to the feature of these state reports observing the same trade repeatedly across time, insofar as the trade has not yet matured or cancelled. To obtain a time series of trading activity, one can merge the trade state reports from both dates into a single dataset and deduplicate that inflation swap transaction which now appears twice. This deduplicated dataset will then have a single inflation swap contract that was executed on date T.[1] This is also the main idea behind our data cleaning procedure.

One major challenge working with trade repository data lies in the sheer volume of observations and the non-universality of data structure across trade repositories and across time. In the DTCC trade repository, for instance, a given trade state report from a *single* trading day can contain between 2.5 to 3.5 million observations. In comparison, a given trade state report from the UnaVista trade repository can contain between 15 to 25 million observations. This is due to the fact that the UnaVista trade repository stores the records of all derivative transactions into a single trade report regardless of their underlying asset class. This differs from the DTCC trade repository, where trade reports are individually generated for each asset class. This implies that considerable computational power and memory is needed for data ingestion.[2] Each trade repository also contains a different number of data fields: trade state reports from DTCC, UnaVista, ICE and Regis-TR have 186, 152, 137 and 138 variables respectively. Some variables e.g., "fixed rate of leg 1", can also sometimes be recorded as a numeric-type variable and sometimes a string-type variable, depending on the trade repository, and these are also not exactly consistent across time. Trade reports generated by the same trade repository itself can also sometimes have a varying number of variables. In our experience working extensively with the DTCC trade repository, we

[1] Such deduplication is also required even if trade activity reports are used instead of trade state reports to construct a daily time series of trading activities. This is because it is not always the case that both counterparties to the transaction reports it to the trade repository at the same time. One counterparty can report it in the evening of date T for which the trade was executed, but the other counterparty can report it only in the afternoon of date $T+1$. This implies that trade activity reports generated on both dates would capture the same derivative transaction each, with the exception that it was reported by a different counterparty. Yet, the transaction is pairwise in the sense that it shares the same pair of counterparties and has the same trade ID. Thus, one should append both trade activity files and deduplicate this pairwise transaction such that only one remains in the dataset.

[2] Our data cleaning procedure below takes roughly 50 minutes to complete for every DTCC trade state report, and approximately 2.5 hours for every UnaVista trade state report.

found that on some random dates in our sample the trade state files are generated with 190 variables — instead of 186 — where the 4 additional variables are all empty. Some DTCC trade state reports also have wrongly reported fields for some transactions: e.g., “XXXX” was input as the notional amount of the transaction where it should have been recorded as the trading venue. It is crucial for any data cleaning infrastructure to be able to handle these ‘special cases’ especially if it is automated by a loop. All these implies that any data-cleaning infrastructure would have to be tailored for *each* trade repository, and quite some cost is required to adapt it such that it is equally capable of handling data from another trade repository.

2.3 Data cleaning procedure

In what follows, we provide a detailed documentation of our data cleaning procedure primarily designed to handle data from the DTCC trade repository. Most of these can also be adapted wholesale to the other trade repositories.

We first define the preliminaries of our empirical work by restricting our data sample to daily DTCC OTC interest rate trade state files from 31^{th} October 2017 to 10^{th} February 2023 since we focus on inflation derivatives. Inflation derivatives are a subset of interest rate derivatives, where the underlying asset of the interest rate derivative is a floating inflation rate index. Thus, our entire raw dataset consists of 1,321 trade state files, each containing a stock of approximately 2.5 to 3.5 million outstanding transaction-level trade reports. The total number of initial observations is approximately 4 billion.

Next, we describe the main steps taken to clean each individual DTCC OTC interest rate trade state report. We then describe the automation that was built to allow for the same code to be implemented on each trade state report at a daily frequency:

1. **Identify inflation derivative contracts from the raw trade state report**. To take advantage of potential speed gains in our cleaning procedure, we begin by identifying inflation derivatives from the entire stock of outstanding contracts and drop the remainder of the dataset. This has the advantage that the remainder of our extensive cleaning procedure can be applied to a smaller subset of observations and is therefore significantly faster.

   To do so, we first extracted the string that is associated with the floating leg fields of the derivative contract and used regular expressions to exhaustively capture strings that refer to inflation indexes. We find that the following strings are sufficient enough to capture the entire universe of inflation derivatives: “cpi”, “rpi”, “ukpri”, “lpi”, “hicp”, “-infla-”, “inflatn”, “inflation”, “inf”, “cpx”, “cpt”, “consume”, “cpunr”, “ukrp”, “ukcp”, “urnsa”, “cpurn”, “non-revised”, “nonrevised”, “non-re”, “harmonised”, “tobacco”, “excluding tobacco".[3] Subsequently, we checked for the product classification type of the derivative contract (the relevant variable is "product classification") and used its ISO 10962 6-character CFI code to verify whether it is an inflation swap contract. For example, the CFI code "SRGCSC" stands for: (S): Swaps; (R): Rates; (G): Underlying assets: inflation rate index; (C): Notional: constant; (S): Single or multi-currency: single; (C): Delivery: Cash. A derivative with "SRGCSC" as its product classification would be indicative that it is an inflation swap. On this premise, we add a secondary condition that the transaction must be recorded with a recognizable inflation rate index — for instance, this would be picked up by the preceding regular expression functions — or else identifying this inflation swap would not be useful insofar as we cannot identify whether it is a UK inflation or US inflation index, etc. We only kept the trade reports that can be identified with a recognizable inflation index.[4] In this part of our cleaning procedure, we coded up the entire ISO-10962 (and all its 4 classes of attributes) which allows us to identify as many as 14 different categories of derivatives.[5] Following this procedure, we can subsequently identify 5 categories of inflation derivatives from any given trade state report. These are: (i) swaps, (ii) listed options, (iii) non-listed and complex-listed options, (iv) strategies and (v) miscellaneous, with inflation swaps accounting for roughly 97.5% to 98.5% of overall number of transactions.

2. **Drop matured trades, terminated trades and forward-starting trades that will only go into effect 10 years later or more**. Next, we shrink the dataset further by dropping the transactions that are not economically important: these are the trades that have either matured, been terminated, or those that only go into effect in the far future after 10 years. We identify these using the information on its valuation date, maturity date, termination date and effective date.[6] We also drop a minority of trades with a valuation date that falls on a weekend since this is not consistent with market convention.

3. **Deduplicate pairwise consistent transaction reports at the counterparty-level**. Owing to UK-EMIR reporting requirements, all UK-regulated counterparties to a derivative transaction are required to report it to a relevant trade repository. This implies that our dataset up to this juncture contains a pool of duplicated transactions that are pairwise, i.e., they share the same trade ID, and have the same pair of LEIs identifying the two counterparties to the trade with one recorded as the "reporting counterparty ID" in one trade report submitted

[3] On the set of derivatives that contain these strings in their floating rate data fields, we further used regular expressions to check whether there are some transactions with "clpicp" as their floating rate index. These derivatives could be present due to the string "lpi", which was meant to capture inflation derivatives with limit price indexation. We then drop these transactions, as "clpicp" refers to CLP-ICP fixed-to-floating interest rate swap (the floating rate refers to the Chilean Average Chamber Index (Índice de Cámara Promedio or "ICP").

[4] The number of inflation derivative transactions reported without a recognizable inflation index (or if any at all) are few, and account for less than 0.2% of the total number of observations at this juncture.

[5] These are: "Equity", "Collective Investment Vehicles", "Debt Instruments", "Entitlements", "Options", "Futures", "Swaps", "Non-listed complex options", "Spot", "Forwards", "Strategies", "Financing", "Reference Instruments", "Miscellaneous".

[6] For example, matured trades are those with a valuation date that precedes their maturity date.

by the reporting counterparty and this very LEI is simultaneously recorded as the "ID of the other counterparty" when the same transaction is reported by the other counterparty. These pairwise observations have to be deduplicated or else there would be double counting in our data when we calculate net and gross positions.

To do so, we first sort the transaction-level observations by its trade ID. Deduplication based solely on trade ID would not be the most precise approach, as it does not uniquely identify the trade. Each trade is uniquely identified only at the *counterparty-level*, that is, the uniqueness of each transaction requires the combination of trade ID and the two LEIs of both counterparties at the very least.[7] For every group of trades with the same trade ID, we carefully identify those that are pairwise and those that are individually unique. This step requires a procedure that is both exhaustive and targeted: this is because we can have groups up to 8 to 10 trades all with the same trade ID, but within the group there may only be 2 or 4 trades that are pairwise, that may either be pairwise with regards to the same pair of LEIs (in which we drop the pair of older reports) or to a set of two pairs of LEIs.[8] We work through this combinatorial problem carefully to arrive at a dataset that has a pool of individually unique trade reports and another pool of pairwise trade reports.

We then focus on the pool of pairwise trade reports to check for internal consistency of the information submitted by both reporting counterparties. We cross-validated each pair of trade reports by the following criteria:

(a) Notional amount: both counterparties should report the same notional amount in the transaction, at least within a rounding margin. We also check that the currency in which the notional amount is reported by both counterparties is identical.[9]

(b) Maturity date: both counterparties should report the same maturity date on the contract.

[7] As an illustration, consider a dataset sorted according to the trade ID of each transaction-level report. All the trades with the same trade ID e.g., 001, will therefore be grouped together. However, one trade may have a pair of counterparties A and B while another has counterparties A and C. This implies that both trades are individually unique and are not a pair of duplicated reports.

[8] Consider the following scenario: suppose we identify a group of 4 trades that share the same trade ID, but there exists two sets of pairwise bundles reported as:

- (i) Trade ID = 001, reporting counterparty ID = A, ID of the other counterparty = B
- (ii) Trade ID = 001, reporting counterparty ID = A, ID of the other counterparty = C
- (iii) Trade ID = 001, reporting counterparty ID = B, ID of the other counterparty = A
- (iv) Trade ID = 001, reporting counterparty ID = C, ID of the other counterparty = A

In this scenario, the trade reports in (i) and (iii) are pairwise while the trade report in (ii) and (iv) are pairwise. Hence, the goal would be to deduplicate one trade from each of these pairs. However, in another scenario we could well have that the ID of the other counterparty in trade (iii) *not* to be A, in which case the trade reports in (i) and (iii) become individually unique and should not be deduplicated.

[9] That is, we do not include cross-currency inflation swaps in our analysis. These are non-standard inflation swaps, and are extremely few in the dataset.

(c) Intragroup flag: both counterparties should report consistently whether the trade is an intragroup transaction.

(d) Counterparty side: each pair of pairwise transactions must consistently indicate whether one LEI is the buyer (and the other a seller) and vice versa.

There is no standard formula for what criteria to consider. We considered these as they are of first-order importance in determining the precision of our calculations of gross and net notional positions. For example, if the pairwise trades are not consistent in the notional amounts reported by both counterparties, we drop the pair altogether because it would not be possible to determine which notional reported is correct. Similarly, we drop the pairs if they are inconsistent with their maturity dates, or else we cannot be precise about the initial time-to-maturity of the contract and thus our market segmentation facts.

It is on this pool of cross-validated pairwise trade reports that we deduplicated each pair and kept only the latest trade report by its reporting timestamp. This allows us to obtain a dataset containing only inflation derivatives transactions that are unique at the counterparty-level. We further drop a minority of these unique trades if they (i) have implausible notional amounts (less than \$1,000 or more than \$10bn); (ii) have a missing counterparty side; (iii) have counterparties that are not identified with the 20-alphanumeric character LEI codes; (iv) are intragroup transactions or compression trades.

4. **Remove reports that do not adhere to UK EMIR Validation Rules**. Next, for every reported derivative transaction we check if the contract information satisfies the set of UK EMIR validation rules provided by the UK's Financial Conduct Authority. These rules list a set of conditions that uniquely apply to each of the 106 variables that can be populated when a transaction is reported, which, depending on whether it is a trade-level or position-level report, is either mandatory (either conditionally or unconditionally) or optional. The conditions listed for each variable may also be interdependent. For example, consider the variables "value of contract", "valuation type" and "cleared". Then, for a cleared transaction with a reported value of contract, its valuation type should be reported as "C" (CCP's valuation) instead of "M" (mark-to-market) or "O" (mark-to-model). We apply these validation rules to the entire dataset consisting of roughly 180,000 outstanding transactions at this juncture, insofar as the reporting timestamp of the transaction is later than 11pm on 31st December 2020.[10] Roughly 5% of reported transactions are in violation of these rules and are removed from the dataset. We note that these validation rules are beneficial for the quality of trade repository data since reporting counterparties are also required to inform the FCA of any breaches of these validations rules. Subsequently, we drop all swap transactions that were not confirmed according to Article 12 of Commission Delegated Regulation No.

[10] This is the exact time when the UK EMIR validation rules become applicable. Stated amendments to these rules apply to transactions reported from 21st June 2021.

149/2013 (link). We take these steps to be conservative towards the data.

5. **Categorizing various inflation markets**. In this step, we return to our regular expressions to properly categorize the reported derivative transactions into their respective markets. This is crucial for completeness as there are generally no universal standards for how an inflation index (tied to the floating rate) should be reported to trade repositories. For instance, we were able to extract strings "ukrpi", "ukpri", "rpi", "ukrp", "GBP - Non-revised Retail" from various swap contracts that should all be categorized as UK RPI inflation swaps. Similarly, swap contracts with floating rate strings such as "ukcpi", "ukcp", "uk-cpi", "GBPNONREVISEDCONSUMERPRIC", "gbcpi", "GBPINFLATNREFB", "GBP-INFLATN-REFB", "gbp cpi", "GBP_CPI", "GBP-CPIUK-INFLFIX" should be categorized as UK CPI inflation swaps.[11] To arrive at an exhaustive list of strings that is truly representative of the inflation products traded in this market, we build and test our code on individual trade state files from many dates, adding more strings to the list when we identify them to have been left out by the list of strings above. We perform this procedure iteratively until no further 'new' strings can be identified pertaining to the relevant inflation index. Using this approach, we carefully identify recognizable inflation rate indexes from 19 different countries/regions in total, with the UK, EU and US inflation markets the largest, and with UK RPI inflation swaps being the most traded derivative product within the UK market.[12]

6. **Allocating counterparty LEIs to an investor group using a best-endeavor sectoral classification**. Given that the LEIs of all counterparties to a derivative transaction are reported as part of the contract information, we can identify these institutions and classify them into an investor group. We use the LEI reported in "beneficiary ID" as opposed to the "reporting counterparty ID", as by definition that entity is the true party subject to the rights and obligations arising from the contract.[13] We then use the "ID of the other counterparty" to identify who is on the other side of a given transaction. This process is naturally subject to errors, like allocating an insurer with asset management arm, so we manually verified and corrected as best as we could.

[11] For EU CPI inflation swaps, the strings we identified are: "EUR EXT CPI", "EUR-EXT-CPI", "EUREXTCPI", "EUR_EXT_CPI", "EURINFLATNREFB", "EUR-INFLATN-REFB", "CPALEMU", "INFLEURNREXTCPI", "INFEUR", "EUR-INFLA", "EUR-EXT-R-CPI", "EUR-CPIEU-INFLFIX", "INFLEUR", "EUCPXTOB", "inflation EUR", "EU CPI XT", "EUR EXCLUDING TOBACCONONR", "EUR CPI", "EXT CPI". For EU HICP inflation swaps, the relevant strings are: "cpxtemu", "CPTFEMU", "CPTFEM", "eu hicp", "eur hicp", "EUROSTATEUROZONEHICPEXTOB", "EUR - Excluding Tobacco-N", "EUR-HICP-REFB", "EURHICPREFBXT05", "EURHICPREFB", "BLGHICP", "EUR-Excluding Tobacco-Non", "EUR-HICPX", "HCPIxt", "EUROZONE HICP", "EUR ZONE HICP EX TOBACCO", "hicpxt", "EUHICPXT", "BLG-HICP", "EUHICP". For US CPI-U inflation swaps, the relevant strings are: "cpurnsa", "cpurns", "usd", "uscpi", "us cpi", "usa", "CPI-U", "cpurn", "INFLUS".

[12] These 19 countries/regions are: Australia, Germany, Spain, Euro Area, France, Israel, Italy, Japan, Sweden, United Kingdom, United States, Ireland, Mexico, Denmark, Norway, Canada, Chile, Switzerland and the Netherlands.

[13] For a overwhelming majority of the transactions, the beneficiary ID exactly coincides with the reporting counterparty ID.

Steps [1]—[6] fully describe the procedure required to clean a single DTCC OTC interest rate trade state file extracted from a given date, which enables us to identify an outstanding stock of approximately 130,000 to 160,000 inflation swap contracts. We then build an automation that can iterate the abovementioned cleaning procedure over each of the 1,321 trade state files from 31^{st} October 2017 to 10^{th} February 2023 at a daily frequency.[14] To conserve computational memory, our automation fetches one raw trade state report from the trade repository locally in each iteration, and deletes it once the cleaning procedure is complete and the cleaned dataset is saved. This entire process took approximately 3 weeks to complete on 5 high-powered servers with 16GB of memory each.

Next, we append these cleaned DTCC OTC interest rate trade state reports into a single dataset. Given the gigantic volume of data, this is computationally feasible only if we were to first reduce the file size of each of these cleaned trade state reports. We do by encoding string-type variables where applicable, dropping irrelevant variables and compressing the data. This is sufficient for appending 62 trade state reports at a monthly frequency between 31^{st} October 2017 to 10^{th} February 2023, which we use to construct a monthly time series of both gross and net notional positions.[15] To construct the daily time series data used to implement our identification strategies, we append 1,321 trade state reports at a daily frequency by further restricting the data sample to the dealer-client segment of the UK RPI inflation swap market. This resulted in a total of 33,784,686 observations in total. We now turn our attention to this dataset and describe the further steps required to collapse it such that it can be used to estimate a VAR.

7. **Removing replica transactions that are repeatedly observed across time**. We begin by first implementing another substantial deduplication procedure to remove repeated observations of the same transaction at the counterparty-level.[16] These repeated observations arise from having the outstanding stocks of executed transactions from daily trade state files merged into a single dataset, and thus the same transactions are repeatedly observed over time (by its execution date) insofar as they have not been terminated or matured. Since the objective is to obtain a time series of trading activities that is reflective of the change

[14] In the DTCC trade repository, it is sometimes the case that two or even three trade state reports are generated for a given business day. They are then archived with a different timestamp. Our procedure extends to all these additional reports, with the caveat that we later implement another round of deduplication when constructing our daily time series to remove identical transactions that are double-recorded by these additional trade state reports.

[15] This procedure additionally requires one to remove some stale trades in the dataset. These are identical trades at the counterparty-level (i.e., have exactly identical trade ID and LEIs for reporting counterparty ID and ID of the other counterparty to the transaction) that have the same valuation dates. These valuation dates should be refreshed based on the recency of the trade state report generated. Thus, we deduplicate these stale trades and keep the latest transaction so as to avoid double counting.

[16] Note that this procedure differs from deduplicating pairwise trade reports in Step [6]. These trade reports are defined as "replicas" because they take the following form in the sorted dataset e.g.,

- (i) Trade ID = 001, reporting counterparty ID = A, ID of the other counterparty = B, execution date = T
- (ii) Trade ID = 001, reporting counterparty ID = A, ID of the other counterparty = B, execution date = T

and can differ in their reporting timestamps and valuation timestamps.

in net notional positions of each counterparty, this deduplication is an essential step. We sort the counterparty-level transactions by its execution date forward in time, and only keep the latest contract reported to the trade repository for a given execution date of the trade.[17] For the replicated trades that all share the same reporting timestamp, we keep the one that has the latest valuation timestamp. For the replicated trades that have identical timestamps for both its reporting date and its valuation date, we consider them to be stale trades and keep one arbitrary report. Upon completion of this step, we also repeat the deduplication of pairwise trade reports that may be observed owing to different trade state reports being combined into a single dataset. All these sequences have to be carefully implemented in order to arrive at an unbalanced panel of 145,181 unique, transaction-level UK RPI swap trade reports. This can now be used to analyse the change in net notional positions: precisely because each transaction is unique, we can use the notional amount quoted as a direct measure of the change in net notional position for the pair of counterparties involved. Since these counterparties have already been classified into an investor type in Step [6], we can further aggregate these positions taken by all actively trading institutions on that particular execution date to calculate the change in net notional position at a sectoral level (i.e., pension funds against dealer banks).

8. **Merging in Bloomberg prices**. Next, we merge our transaction-level data with another dataset containing prices obtained from Bloomberg by the execution date of each transaction.[18] These are daily UK RPI inflation swap rates for zero coupon swap contracts with an initial time-to-maturity of 1, 2,..., 10, 12, 15, 20, 25, 30, 35, 40, 45, and 50 years. To match the relevant prices to every contract, we calculate the initial time-to-maturity for the entire pool of transactions using their maturity dates and effective dates.[19] It is then straightforward to match these prices to the trades that have an initial maturity that exactly matches those from the zero coupon contracts for which the Bloomberg swap rates are priced for. We focus our attention on the trades for which their initial maturity is not a whole integer (when measured by number of years).[20] For these contracts, we implement a "nearest neighbour" method. That is, an inflation swap with initial maturity of 2.8 years will be regarded as a 3-year swap and be matched with a 3-year swap rate, while a swap with initial maturity of 13 years will be regarded as a 12-year swap and be matched with a 12-year swap rate, and so on. In particular, those with an initial maturity longer than 30-years will all be matched

[17]This step is taken as a proxy for the report to contain the most updated information or terms of trade. Note that while these replica transactions reported to the trade repository can have a different reporting times, by definition they all have the same execution date since they ultimately refer to the same trade.

[18]These swap prices have been adjusted for RPI indexation lags. See Appendix 1.1 for a formal description of the problem.

[19]We also drop a minority of trades in this step for which we are unable to calculate the initial maturity, either because the maturity date or effective date fields are not populated.

[20]For instance, we are able to observe an inflation swap with an initial time-to-maturity of 2.7 years. These contracts are to be expected, due to the highly OTC nature of the market that allows terms of trade to be customised.

with the 30-year swap rate.

9. **Constructing daily price-quantity pairs in the long and short maturity markets**. This procedure transforms the transaction-level dataset—which features trade-by-trade transactions at a counterparty-level sorted by an execution timestamp precise to the seconds—into a daily time series featuring a pair of prices and quantities for both the short maturity and long maturity markets. Further, it restricts the data sample to dealer-client trades in UK RPI inflation swaps where the client is either a hedge fund or pension fund (this also includes the LDI funds), and all trades with an initial maturity that is between 3 years to 10 years are dropped since this segment of the market is not our focus. This yields the segmented market, with hedge funds active in their trading of swaps with initial maturity 3 years or less and pension funds primarily trading in the long maturity market. Since the objective is to construct one price-quantity pair for each of these markets, and there are multiple prices within each market (e.g., since the short maturity market consists of all swaps with an initial maturity of 3 years or less, it has a composition of 1-year, 2-year and 3-year swap rates), we remove such compositional effects by weighting these prices by their gross notional shares. These shares are calculated from hedge fund's trading activities throughout the entire data sample. We replicate this procedure for the long maturity inflation market, except that prices in this market are weighted by the gross notional shares of the corresponding maturities traded by pension funds. This ensures that there is just one price corresponding to each execution date, that is also identical across contracts with different initial maturities insofar as they belong to the same market. We subsequently aggregated up all notional positions at a transaction-level (taking into consideration whether it is a contract bought or sold to the dealer) in both markets transacted within a given execution date and collapsed the data by scaling up the unit of analysis from a second to a day.

3 Data: additional information

This section of the Appendix provides additional information on the data used in the empirical analysis. Section 3.1 defines the key variables used in our VAR implementation. Section 3.2 compares our data with data from a different source.

3.1 Definition of the key variables

The DTCC trade repository data gives us positions on the inflation swap market at each day, t, for institution i defined by its legal entity identifiers, across the three main sectors b, h, f as well as others, and for the maturity of the contracts per year. We aggregate contract maturities into three buckets: three years or less into P, Q, and 10 years or more into p, q, leaving out the remainder. Noting that data quality is poor prior to 2019, we dropped all the corresponding quantities and prices with an execution date that precedes 1st January 2019. This yields a time series dataset from the 2^{nd} January 2019 to the 10^{th} February 2023, for 1078 days. Therefore, we have 1078 x 4 observations in the raw data matrix used for VAR estimation in implementing all of our three identification strategies.

The two key observables used in estimation are a balanced panel for $q_{f,i,t}$ and $a_{f,i,t}$ (and same for b and h). The trading activity corresponds to $q_{f,i,t}$, while the $a_{f,i,t}$ are the gross notional amount of all outstanding inflation swap contracts traded by institution i from the pension fund sector. To build this measure, we carefully tracked the trading activity of each institution in our data sample and accumulated the stock of its outstanding positions by taking account of not only new inflation swap trades, but also trades entered into at the earlier part of the data sample that have eventually matured prior to the cessation of our data sample.

Every contract has a separate price and each maturity has a different price within the long and short buckets. We build the market price p_t as the weighted-average daily price of a UK RPI zero coupon inflation swaps of initial time-to-maturity 10 years or more, where the weights are gross notionals traded in each long maturity category by pension funds as a share of the total across the data sample. Likewise P_t is the weighted-average daily price of UK RPI zero coupon inflation swaps with weights equal to the share of gross notional amount traded in each maturity category by hedge fund institutions in this market.

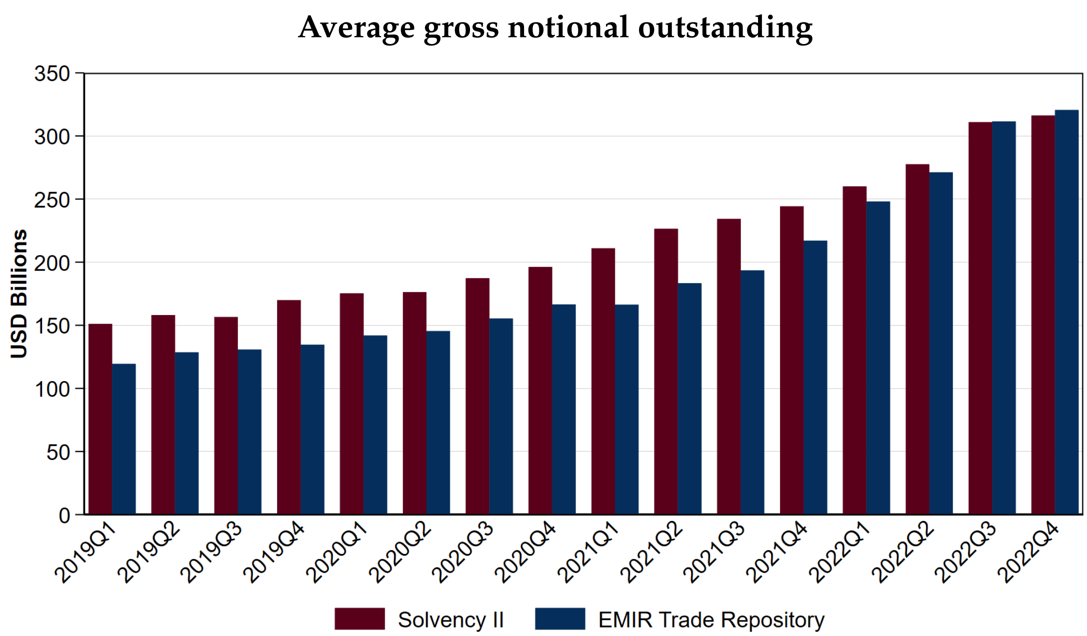
3.2 A robustness check: comparison to alternative data

To check that our trade repository dataset is representative of the trading activity in the OTC inflation swap market, we collected supervisory data on the derivative holdings of insurance companies that are regulated by the UK's Prudential Regulation Authority (PRA) and subject to the Solvency II Directive. Most insurers within scope of the Solvency II Directive are required to submit annual and quarterly returns, with the exception of some smaller firms with quarterly waivers. The reports include detailed information on the derivatives holdings of a given insurer, including the identity of the counterparty, the underlying security, the notional amount, and the derivative category (e.g., inflation swap). Given the supervisory nature of the reporting, we can assume that the Solvency II data provide an exact quarterly snapshot of the total inflation swap holdings in the insurance sector.

Figure 2 compares the average gross notional outstanding of the insurance sector in our dataset with the supervisory holdings in the Solvency II data for the period 2019 Q1 - 2022 Q4. The figure shows that the EMIR TR data cover the vast majority of trading activity in the OTC inflation swap market as reported to the Solvency II database. In 2022 Q4, for example, both datasets report a gross notional of around $320bn for the UK insurance sector.[21]

Throughout our sample period, the EMIR cleaned trade repository data of trading activity in the OTC inflation swap market covers 90% of the total inflation swap holdings reported to the Solvency II database. The improved coverage in the second half of our sample is likely due to the increased precision of the regulatory reporting in the EMIR TR data.

Figure 2 COMPARISON OF SOLVENCY II INSURANCE HOLDINGS AND EMIR TR DATA



NOTE: By the Solvency II Directive, most insurance companies within its supervisory outreach are required to submit annual and quarterly returns, and these include detailed information on the derivatives holdings. Given this alternative dataset on the total inflation swap holdings of insurers in the UK RPI market, we compare the total gross notional position of these insurers across our two data sets in this figure to seek an affirmation that our trade repository data cleaning procedure is robust and fit for the task. SOURCE: Solvency II data and DTCC Trade Repository.

[21]Section 4.2 describes the mechanical nature of insurers' inflation swap trading in recent periods in more detail. Our baseline results remain stable when including insurers' trading volumes (section 6.5).

4 Additional descriptive statistics

This section of the appendix provides more facts about the UK RPI swap market.

Section 4.1 provides more data and information on the gross notional positions in the dealer-client segment of the market to complement the net positions in the main text, splitting the data by maturity segments and also investor-types extending beyond just pension funds and hedge funds i.e., the remainder of the market.

We excluded insurers from our institutions. Section 4.2 discusses their features.

Appendix 4.3 provide additional summary statistics for a single trade state report from the DTCC trade repository, taken on the last day of our data sample that is 10*th* February 2023. These statistics cover all outstanding trades in the dealer-client segment of the UK RPI market that are effective as of this date.

4.1 More on gross positions in the dealer-client inflation swap market

Figure 3 digs deeper into gross notionals, splitting them by client sector, maturities, and type of institutions. Gross notionals have grown rapidly, reaching a peak of around $1.1tn in late 2022, with the larger share in contracts for inflation for a long maturity of ten years or more. Across clients, hedge funds have steadily increased their notionals since the COVID-19 market turmoil in 2020 and the reappearance of inflation in 2021, from less than $50bn in 2019 to around $200bn in 2022.

Figure 4 does a double-sort splitting of the gross notional holdings of the three sectors in our analysis by maturity and across time. The segmentation of the market is clear also in gross holdings, even though it is starker on net holdings.

Figure 5 shows the gross notional positions of all other investor types in the market. For completeness, Figure 6 show the equivalent values in terms of net notional positions.

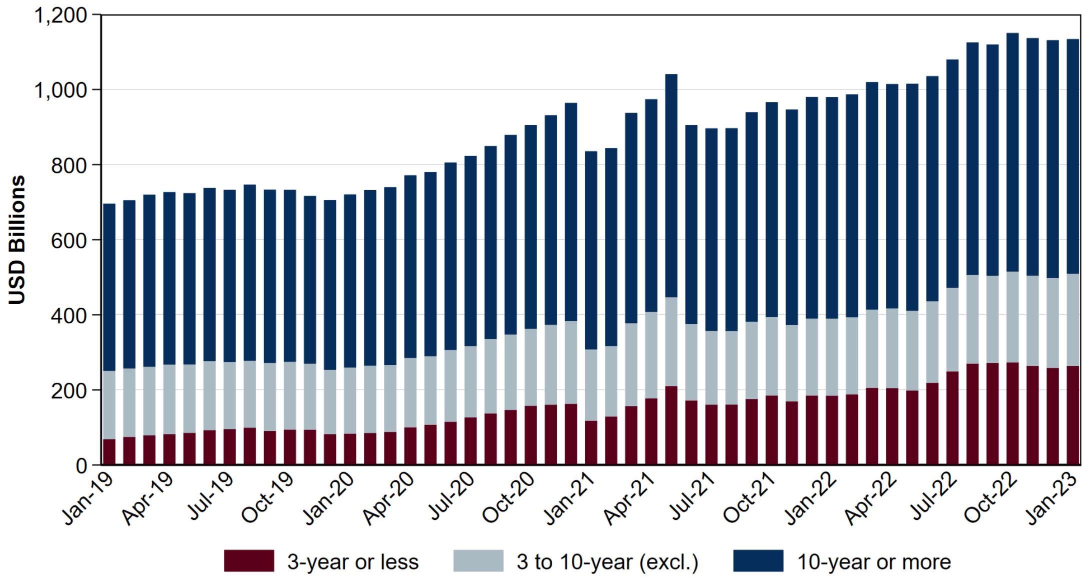
Figure 7 shows the net notional position of dealers broken down by the initial time-to-maturity of the contracts. This is a mirror image of the figure on fact 2 in the main text since dealers are the counterparties of pension funds in the transactions. This also confirms that the transactions with all other clients are not quantitatively significant.

4.2 Insurance companies

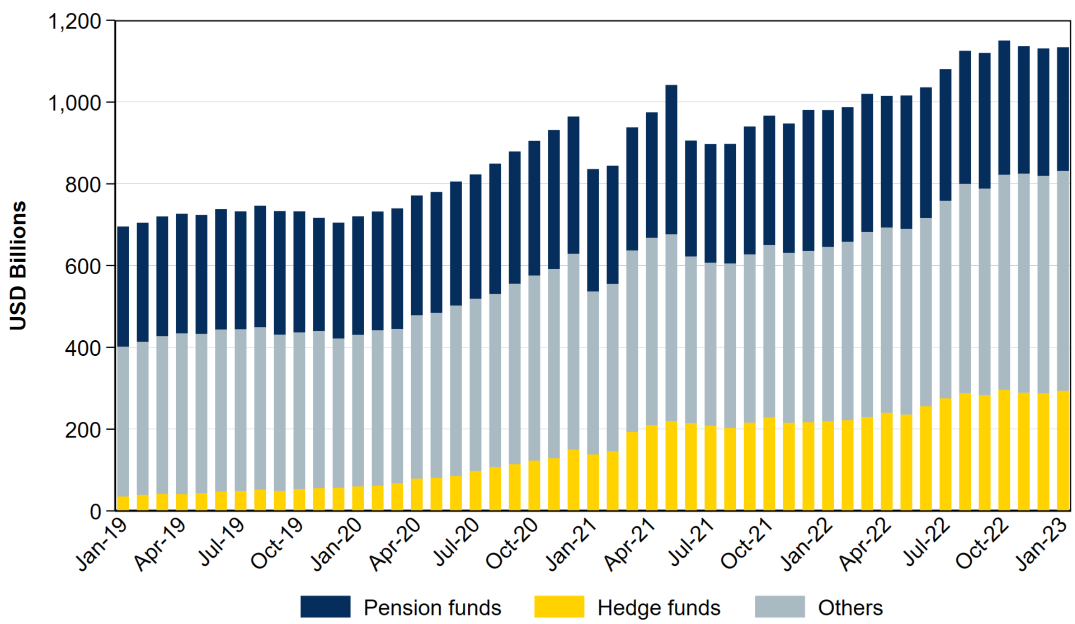
In our baseline results, we focused on the trading volumes of hedge funds in the short-maturity market and pension funds in the long-maturity market. However, as discussed in the main text, there are other types of agents in the market. As noted in Section 3.2, the next notable sector in the inflation swap market are insurance companies, in particular pension insurers. The latter have been heavily involved in the buy-in/buy-outs of pension fund liabilities in recent years, a trend that is set to continue in the coming years.

Figure 3 GROSS NOTIONAL POSITIONS IN THE DEALER-CLIENT SEGMENT

(a) Gross notional by initial time-to-maturity

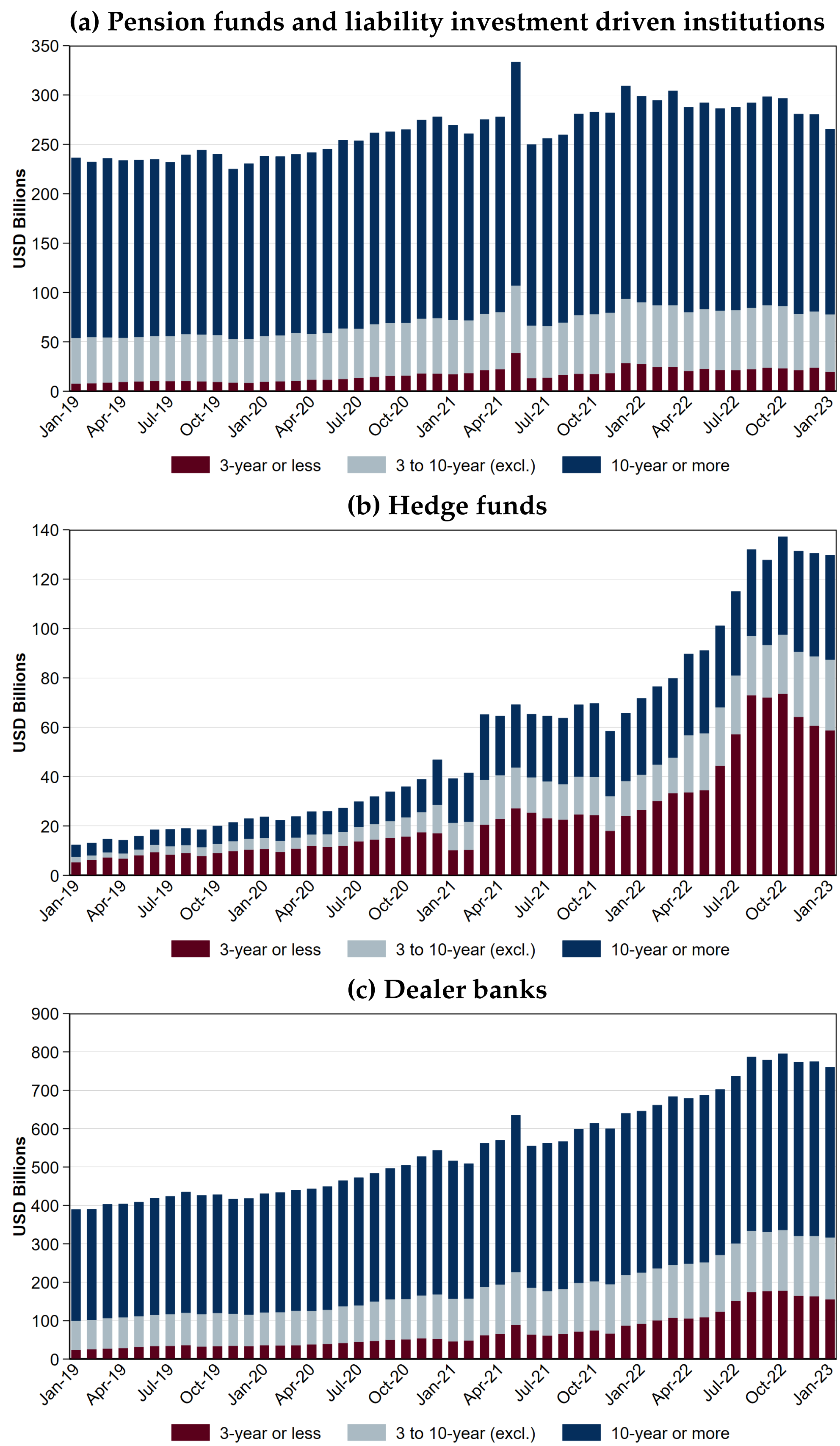


(b) Gross notional by client sector



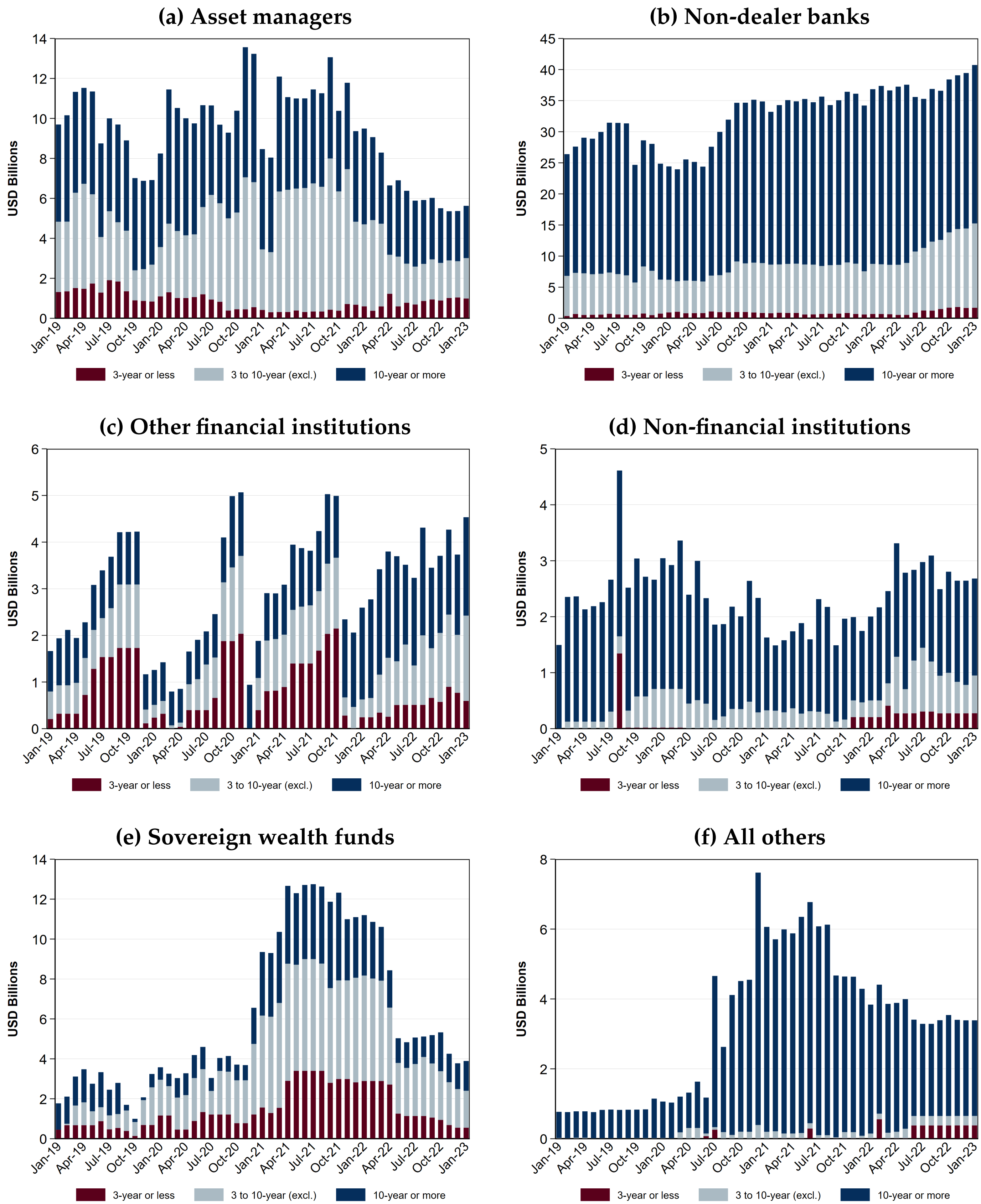
NOTE: This figure shows the gross notional positions outstanding traded in the dealer-client segment of the UK RPI inflation swap market at month-end, categorised by different maturity segments and investor types. These positions represent the gross notional values used to calculate cash flows for both counterparties to the swap contract. Time variation comes from both new contracts executed and the expiration of existing contracts. We omit trades that involve a central counterparty clearing house because the trades net out to zero by construction. The total value for each month is aggregated based on all transaction-level trades with a valuation date-time within the month. The data sample is from January 2019 to February 2023. SOURCE: DTCC Trade Repository.

Figure 4 GROSS NOTIONAL POSITIONS: THE MAIN INVESTORS IN THE UK RPI MARKET



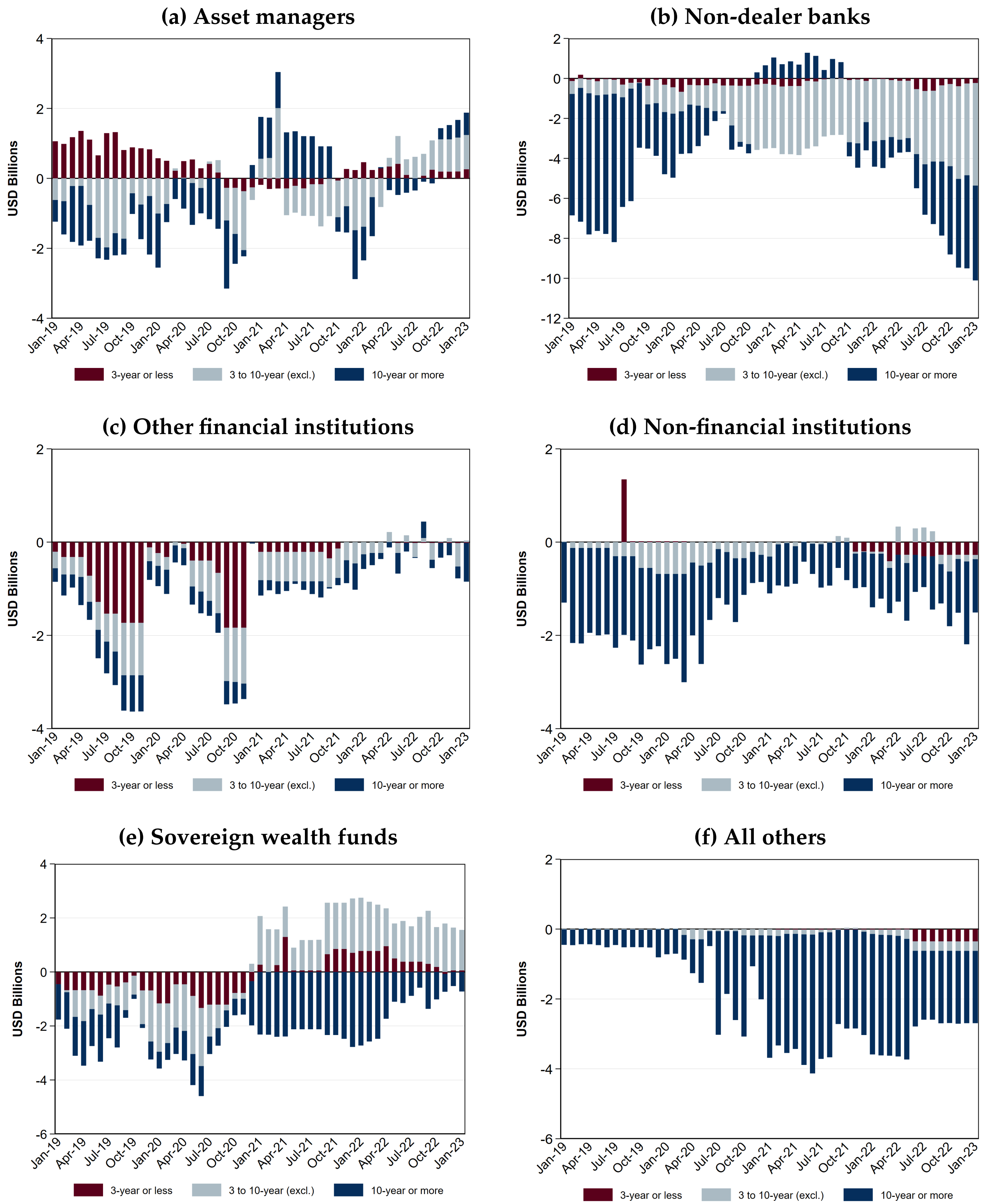
NOTE: This figure shows the gross notional positions outstanding of pension funds, hedge funds and dealers traded in the dealer-client segment of the UK RPI inflation swap market at month-end, categorized by different maturity segments. The gross notional position of dealer banks include their trades with the entire market, not only pension funds and hedge funds. The total value for each month is aggregated based on all transaction-level trades with a valuation date-time within the month. The data sample is from January 2019 to February 2023. SOURCE: DTCC Trade Repository.

Figure 5 GROSS NOTIONAL POSITIONS: THE REMAINDER OF THE UK RPI MARKET



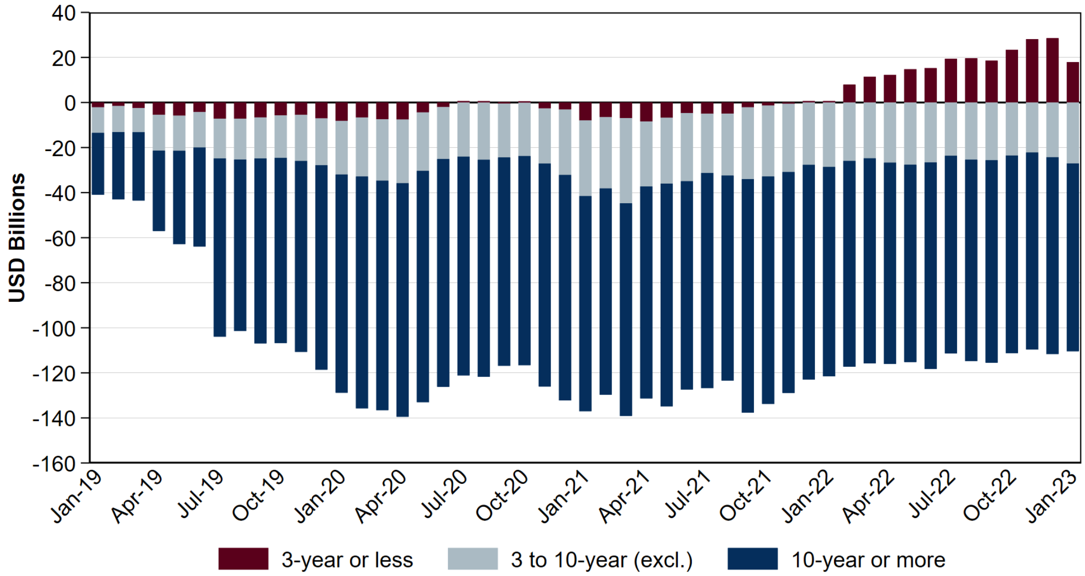
NOTE: "All others" include: state, supranational, proprietary trading firms, trading services and central banks. SOURCE: DTCC Trade Repository.

Figure 6 NET NOTIONAL POSITIONS: THE REMAINDER OF THE UK RPI MARKET



NOTE: "All others" include: state, supranational, proprietary trading firms, trading services and central banks. SOURCE: DTCC Trade Repository.

Figure 7 NET NOTIONAL POSITION OF DEALER BANKS IN THE UK RPI MARKET



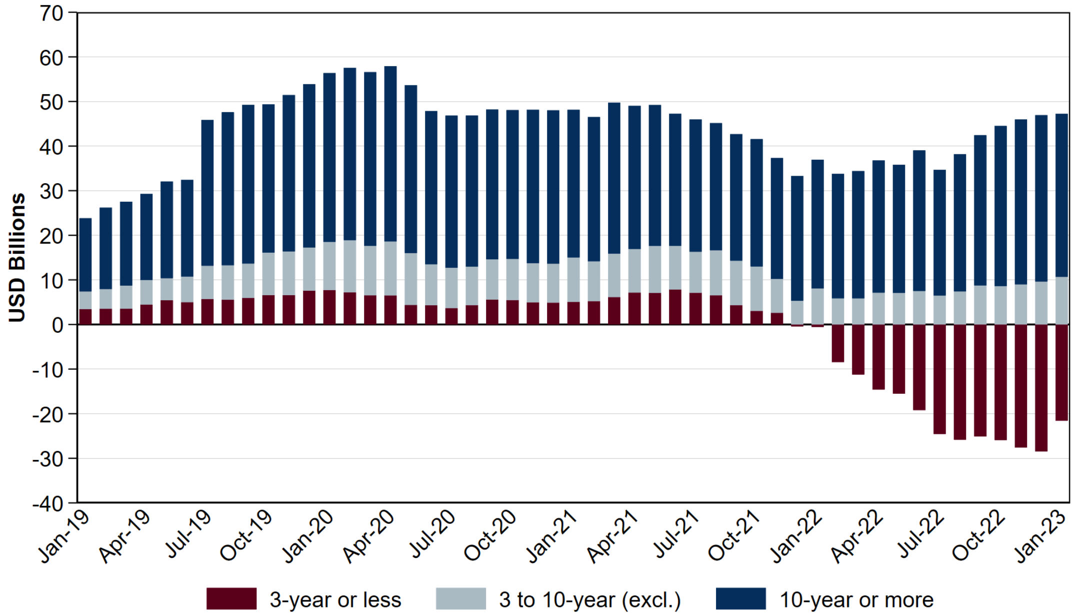
NOTE: This figure shows the aggregate net position of dealer banks in the UK RPI inflation swap market at month-end from January 2019 to February 2023. These positions are an aggregate across the net positions of the largest G17 banks that are licensed to deal in the UK financial market, and they are also an aggregate of all net positions taken against all the abovementioned types of different investor-sectors, with pension funds and insurers holding the majority of the opposite position (See Figure 2 in the main text). A negative net position indicates that dealer banks pay floating inflation to the counterparty, bearing the risk that inflation goes above the contracted swap rate. The total value for each month is aggregated based on all transaction-level trades with a valuation date-time within the month. SOURCE: DTCC Trade Repository.

Including insurers in our analysis is difficult for one main reason. In contrast to pension funds, insurers mainly use cash-flow driven investment strategies (CDI). The aim of a CDI strategy is to create an asset/derivative portfolio that closely matches the cash-flows on the liability side. In terms of the inflation-indexation of pension liabilities, the most prevalent form sees inflation linkage floored at zero and capped at 5%, as measured by RPI.

For a CDI-investor that is fully inflation hedged, inflation moving above the 5% is usually a positive outcome: while the investor's inflation-linked asset rises in value, the liability stops tracking the higher inflation and effectively becomes a nominal liability for a period. The insurer will see its assets rise in value by more than its liabilities. This creates a hedging mismatch: the fund now has too much inflation-linkage. Given that RPI inflation has been above the 5% cap since late 2021, pension insurers have become net sellers of short-dated inflation swaps to reduce their over-hedged positions: see Figure 8.

This re-balancing is a mechanical consequence of the CDI. Their trading activity is passive, and does not depend on prices and beliefs. At an extreme, if it was completely mechanical, insurers would be infra-marginal in the market. Therefore, excluding insurers' trading volumes, as we did in our main analysis, would have no effect on our results. Still, to make sure, we checked that our baseline results remain robust to the inclusion of insurers' quantities.

Figure 8 NET NOTIONAL POSITION OF INSURERS



NOTE: The figure shows the aggregate net notional position of insurance companies *vis-a-vis* dealers in the UK inflation swap market at month-end from January 2019 to February 2023. A positive value indicates that these insurers are net buyers of inflation protection. The total is categorised by the maturity of the underlying swap contract. SOURCE: DTCC Trade Repository.

4.3 Summary statistics on individual positions at one data

To give a sense of the micro data behind our estimates, we provide additional summary statistics for a single trade state report, covering all outstanding trades in dealer-client segment of the UK RPI inflation swap market on the last day in our sample: 10^{th} February 2023.

Table 1 shows that we observe the largest number of trades by pension funds, followed by hedge funds and non-dealer banks. The largest average trades size are for hedge funds with an average notional of $70m. When analyzing the notionals across maturities, we find the largest outstanding notionals for the 1-year, 10-year, and 3-year contracts. The average trade size seems to be decreasing with the maturity of the contract.

Table 1 SUMMARY STATISTICS: DTCC DERIVATIVES REPOSITORY TRADE STATE REPORT

	Gross Notional	Mean	Std. Deviation	5th Percentile	25th Percentile	50th Percentile	75th Percentile	95th Percentile	Total Observations
By investor type									
Pension funds	296,223.7	21.6	(43.9)	0.5	2.8	8.4	23.2	84.5	13,696
Hedge Funds	162,600.3	69.4	(111.3)	5.9	15.6	32.8	81.4	236.3	2,342
Non-dealer banks	44,491.2	21.3	(40.1)	0.3	1.8	8.6	24.7	82.7	2,085
Others	341,242.8	37.6	(91.5)	0.4	3.6	13.4	38.5	140.9	9,073
By initial maturity									
3-year or less	186,482.5	89.4	(169.2)	3.6	16.7	46.3	96.8	311.2	2,086
3 to 10-year, excl.	191,680.2	54.7	(88.9)	1.3	7.8	26.0	66.7	204.8	3,505
10-year or more	466,395.3	21.6	(43.9)	0.4	2.7	8.8	23.3	81.3	21,605
All	844,558.0	31.1	(71.7)	0.5	3.4	11.3	30.9	120.2	27,196

NOTE: Units are USD millions for all columns, except the last. The category "Others" includes asset managers, central banks, insurers, non-financials, other financials, sovereign wealth funds, state entities, supranationals, trading services and proprietary trading firms. "All" also coincides with the statistics pertaining to the dealer bank sector since we report statistics on the dealer-client segment of the market. SOURCE: DTCC Trade Repository, 10^{th} February 2023 trade state report.

5 Samplers used for Bayesian Inference

In this section, we provide a discussion of the Bayesian samplers implemented across the three identification strategies used in the article. Section 5.2 provides details of the sampler used to implement the heteroskedasticity-based identification, while Section 5.3 outlines the modified sampler that we adopt from Bahaj (2020) for the GIV-based identification strategy.[22] In Section 5.4, we provide a summary of the algorithm adopted from Arias et al. (2018) to implement sign and zero restrictions. Across all three identification strategies, we clearly state the number of draws and effective sampling sizes from which we conduct inference.

5.1 The VAR model in reduced-form orthogonal representation

Given that all three identification strategies rely on the same vector autoregression (VAR) model, we first present the reduced-form representation that is estimated by each of the three Bayesian samplers and set the general notation, before elaborating on each sampler in the subsections below. The reduced-form VAR specification is given by:

$$\mathbf{Y}_t = \mathbf{c} + \sum_{\ell=1}^{p} \mathbf{\Phi}_\ell \mathbf{Y}_{t-\ell} + \mathbf{u}_t, \ \ \mathbf{u}_t = \mathbf{\Psi}\boldsymbol{\varepsilon}_t \tag{4}$$

where $\mathbf{u}_t = \mathbf{\Psi}\boldsymbol{\varepsilon}_t$ defines its reduced-form orthogonal representation. $\mathbf{Y}_t = [Q_t, P_t, q_t, p_t]'$ is the 4×1 endogenous variable vector consisting of net purchases of inflation swaps from dealers in the short- and long-maturity markets $\{Q_t, q_t\}$ and their respective swap breakeven rates $\{P_t, p_t\}$ at time t (following this order). $\mathbf{c}$ is a 4×1 vector of deterministic constants, and $\mathbf{u}_t$ is a 4×1 vector of reduced-form VAR forecast errors. The reduced-form coefficients are the autocorrelated lag coefficients $\{\mathbf{\Phi}_\ell\}$ and the variance-covariance matrix $\mathbf{\Sigma} = \mathbb{E}[\mathbf{u}_t\mathbf{u}_t']$. The matrix $\mathbf{\Psi}$ identifies the structural shocks $\boldsymbol{\varepsilon}_t$. Across all three identification strategies, this reduced-form representation is estimated with $p = 3$ lags.

Notation. For the purpose of this appendix section, variables and symbols in bold will be used to denote vectors and matrices, and scalars are denoted otherwise without bold-faced symbols.

5.2 Bayesian Structural VAR with Heteroskedasticity-Based Identification

Following the reduced-form representation of the VAR model in Equation (4), the reduced-form orthogonal structural representation of the VAR model under the heteroskedasticity-based identification strategy is:

$$\begin{gathered} \mathbf{u}_t = I_t \cdot \left[\mathbf{\Psi}\boldsymbol{\epsilon}_{H,t}\right] + (1 - I_t) \cdot \left[\mathbf{\Psi}\boldsymbol{\epsilon}_{L,t}\right], \\ \boldsymbol{\epsilon}_{H,t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_H), \ \ \boldsymbol{\epsilon}_{L,t} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Omega}_L), \end{gathered} \tag{5}$$

[22] Michele Piffer provided this sampler for the heteroskedasticity-based identification strategy.

where I_t is an indicator variable denoting the type of regime, with $I_t = 1$ for the high-volatility regime (H) and $I_t = 0$ for the low-volatility regime (L). In our application, the high-volatility regime corresponds to the dates in data sample where there are news releases concerning inflation. $\mathbf{Y}_t$ is a 4×1 vector of 4 endogenous variables. $\mathbf{\Psi}$ is a 4×4 structural impact matrix. The coefficient matrix $\mathbf{\Phi}_\ell$ has dimension 4×4. The residuals $\mathbf{u}_t = I_t \cdot [\mathbf{\Psi}\boldsymbol{\epsilon}_{H,t}] + (1 - I_t) \cdot [\mathbf{\Psi}\boldsymbol{\epsilon}_{L,t}]$ follow a normal distribution, $\mathbf{u}_t \sim N(\mathbf{0}, \mathbf{\Sigma}_t)$, with:

$$\mathbf{\Sigma}_t = I_t \cdot \mathbf{\Sigma}_H + (1 - I_t) \cdot \mathbf{\Sigma}_L, \tag{6}$$

$$\mathbf{\Sigma}_H = \mathbf{\Psi}\mathbf{\Omega}_H\mathbf{\Psi}', \tag{7}$$

$$\mathbf{\Sigma}_L = \mathbf{\Psi}\mathbf{\Omega}_L\mathbf{\Psi}', \tag{8}$$

where $\mathbf{\Sigma}_t$, $\mathbf{\Sigma}_L$, and $\mathbf{\Sigma}_H$ are 4×4 covariance matrices. The structural shocks $\boldsymbol{\epsilon}_{L,t}$ and $\boldsymbol{\epsilon}_{H,t}$ are 4×1 vectors, with $\mathbf{\Omega}_L$ and $\mathbf{\Omega}_H$ as their corresponding 4×4 covariance matrices.

5.2.1 Prior Specification

The prior for the coefficients $[\mathbf{c}, \mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Phi}_3]$ is a multivariate normal distribution:

$$\text{vec}([\mathbf{c}, \mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Phi}_3]) \sim \mathcal{N}(\boldsymbol{\mu}_{\text{prior}}, \mathbf{V}_{\text{prior}}), \tag{9}$$

where $vec(\cdot)$ represents the vectorisation of a matrix. We use uninformative priors imposing $\boldsymbol{\mu}_{\text{prior}} = \mathbf{0}_{\mathbf{52}\times\mathbf{1}}$ and $\mathbf{V}_{\text{prior}}^{-1} = \mathbf{0}_{\mathbf{52}\times\mathbf{52}}$ and let the data inform the posterior distribution entirely. The prior for the structural shock covariance matrices $\mathbf{\Omega}_L$ and $\mathbf{\Omega}_H$ is an inverse-Wishart distribution given by:

$$\mathbf{\Omega}_j \sim \mathcal{IW}(\mathbf{S}, d), \quad j \in \{L, H\}, \tag{10}$$

where $\mathbf{S}$ is the prior scale matrix, and d is the degrees of freedom equal to 6. We impose $\mathbf{S} = \mathbf{0}_{\mathbf{4}\times\mathbf{4}}$ as an uninformative prior and let the data inform the posterior scale matrix entirely, where the posterior scale matrix is positive definite.

5.2.2 Posterior Conditional Distributions

Given the prior specification, the posterior distribution of $[\mathbf{c}, \mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Phi}_3]$ is a multivariate normal distribution:

$$\text{vec}([\mathbf{c}, \mathbf{\Phi}_1, \mathbf{\Phi}_2, \mathbf{\Phi}_3]) \mid \mathbf{Y}, \mathbf{X}, \mathbf{\Sigma}_t \sim \mathcal{N}(\boldsymbol{\mu}_{\text{posterior}}, \mathbf{V}_{\text{posterior}}), \tag{11}$$

where its posterior mean $\boldsymbol{\mu}_{\text{posterior}}$ and variances $\mathbf{V}_{\text{posterior}}$ are defined and updated according to:

$$\mathbf{V}_{\text{posterior}} = \Big[\mathbf{V}_{\text{prior}}^{-1} + \sum_j \mathbf{X}_j \mathbf{X}_j' \otimes \boldsymbol{\Sigma}_j^{-1}\Big]^{-1}, \tag{12}$$

$$\boldsymbol{\mu}_{\text{posterior}} = \mathbf{V}_{\text{posterior}}\Big[\mathbf{V}_{\text{prior}}^{-1}\boldsymbol{\mu}_{\text{prior}} + \sum_j (\mathbf{X}_j \otimes \boldsymbol{\Sigma}_j^{-1})\tilde{\mathbf{y}}_j\Big]. \tag{13}$$

$\mathbf{Y}_j = [\mathbf{Y}_1, ..., \mathbf{Y}_t, ..., \mathbf{Y}_{T_j}]$ is a $4 \times T_j$ matrix of endogenous variables in regime j, where each column corresponds to $\mathbf{Y}_t$ for observations belonging to regime j, and $\mathbf{X}_j = [\mathbf{X}_1, ..., \mathbf{X}_t, ..., \mathbf{X}_{T_j}]$ is a $13 \times T_j$ matrix of regressors in regime j, where $\mathbf{X}_t = [1, \mathbf{Y}_{t-1}, \mathbf{Y}_{t-2}, \mathbf{Y}_{t-3}]'$, with T_j representing the number of observations in regime j and $j \in \{L, H\}$. The notations $\mathbf{Y}$ and $\mathbf{X}$ in Equation (11) are therefore to be interpreted as the collection of $\mathbf{Y}_j$ and $\mathbf{X}_j$, for $j \in \{L, H\}$ such that they recover the raw data sample. The vector $\tilde{\mathbf{y}}_j = \text{vec}(\mathbf{Y}_j)$ is the column-stacked version of $\mathbf{Y}_j$, with dimension $4T_j \times 1$. The notation $\otimes$ refers to the Kronecker product operator.

The posterior distribution of $\boldsymbol{\Sigma}_j$ for $j \in \{L, H\}$ is an inverse-Wishart distribution. Given the residuals $\mathbf{u}$ and a draw of $[\hat{\mathbf{c}}, \hat{\boldsymbol{\Phi}}_1, \hat{\boldsymbol{\Phi}}_2, \hat{\boldsymbol{\Phi}}_3]$ from the posterior distribution of $[\mathbf{c}, \boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, \boldsymbol{\Phi}_3]$, the posterior is $\boldsymbol{\Sigma}_j \sim \text{IW}(\mathbf{S}_j^*, d_j^*)$, where:

$$\mathbf{S}_j^* = \mathbf{S} + (\mathbf{Y}_j - \hat{\mathbf{c}} - \sum_{\ell=1}^{3} \boldsymbol{\Phi}_\ell \mathbf{Y}_{t-\ell})(\mathbf{Y}_j - \hat{\mathbf{c}} - \sum_{\ell=1}^{3} \boldsymbol{\Phi}_\ell \mathbf{Y}_{t-\ell})', \tag{14}$$

$$d_j^* = d + T_j, \tag{15}$$

with T_j representing the number of observations in regime j.

For inference of the inflation shock, we implement a eigenvalue decomposition $\boldsymbol{\Sigma}_L^{-1}\boldsymbol{\Sigma}_H = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{-1}$, where $\mathbf{Q}$ is a 4×4 matrix contains the eigenvectors and $\boldsymbol{\Lambda}$ is a 4×4 diagonal matrix of eigenvalues. The inflation shock corresponds to the eigenvector associated with the largest eigenvalue. We simulate the posterior distribution via MCMC using 1,500,000 draws from the Gibbs sampler, with the first 500,000 draws discarded as burn-in. The remaining chain is thinned by a factor of 100, leaving 10,000 draws for inference. The results are subsequently presented based on the median of these retained draws.

5.3 Bayesian Structural VAR with Granular Instrument Variables

5.3.1 Shock Identification

In this section, we provide further details on the Bayesian implementation of our SVAR using granular instruments as proxies for the underlying demand and supply shocks. We use a modified version of the sampler from Bahaj (2020).

To start, it is useful to first clarify the identification of these shocks and we proceed in a two-step procedure. Recall that the relevance assumption requires:

$$\mathbb{E}(GIV_{f,t}\varepsilon_{f,t}) \neq 0 \text{ and } \mathbb{E}(GIV_{b,t}\varepsilon_{b,t}) \neq 0 \text{ and } \mathbb{E}(GIV_{h,t}\varepsilon_{h,t}) \neq 0, \tag{16}$$

where $\varepsilon_{f,t}$, $\varepsilon_{h,t}$ and $\varepsilon_{b,t}$ are the demand and supply shocks from pension funds, hedge funds and dealer banks respectively, with $\{GIV_{x,t}\}_{x=f,h,b}$ representing their respective granular instruments constructed from the estimated interactive fixed effects factor panel as discussed in the main text. A straightforward representation of the relevance assumption is the imposition that the granular instruments measure the true, unobserved demand and supply shocks with some noise. Such a measurement equation is given by:

$$GIV_{x,t} = v_x \varepsilon_{x,t} + \iota_{x,t}, \tag{17}$$

where v_x is an unknown scale parameter and $\iota_{x,t}$ is assumed to be a measurement disturbance that is orthogonal to the demand and supply shocks. Using matrix notation, we define $\mathbf{M} = [\mathbf{GIV_f}, \mathbf{GIV_h}, \mathbf{GIV_b}]$ as the $T \times 3$ matrix of granular instruments for the reduced-form residual $\mathbf{U}$ (of dimension $T \times 4$), which vertically stacks the residual vectors $\mathbf{u}'_t$, each of dimension 1×4, that we can estimate by least squares, and T denotes the number of total observations in the VAR used for estimation.

Let $\mathbf{A}$ be defined such that it is a 4×3 submatrix of $\mathbf{\Upsilon}^{-1}$ from the VAR specification containing the three columns from the left that identify these demand and supply shocks in sequence (hence ordering the inflation shock last). The structural-form VAR can be represented by:

$$\mathbf{UA} = \boldsymbol{\varepsilon}, \;\; \boldsymbol{\varepsilon} = [\boldsymbol{\varepsilon}_f, \boldsymbol{\varepsilon}_h, \boldsymbol{\varepsilon}_b]. \tag{18}$$

Substituting this structural-form representation into the identification equation above yields the first stage regression that can be estimated by least squares:

$$\mathbf{M} = \mathbf{UY} + \mathbf{V}, \tag{19}$$

where $\mathbf{Y} = \mathbf{A}\boldsymbol{\theta}$ identifies the demand and supply shocks up to a scaling matrix $\boldsymbol{\theta} = diag\{v_f, v_h, v_b\}$ that is diagonal, and $\mathbf{V}$ stacks the measurement errors such that its variance-covariance matrix is given by $\mathbf{\Sigma}_V$. Equation (19) is the first-stage regression that identifies the demand and supply shocks, given the validity and relevance of the granular instruments $\mathbf{M}$.

Once the demand and supply shocks are fully identified from this first stage, we use them to identify the residual shock remaining from a "second-stage" regression that has the interpretation of a inflation shock within the VAR framework. In particular, let $\mathbf{a}_\pi$ be a column vector that identifies the inflation shock i.e., $\mathbf{Ua}_\pi = \boldsymbol{\varepsilon}_\pi$. This vector can be identified by using the demand and supply shocks as instruments to project the first three reduced-form residuals on the fourth, and the residual from this least squares regression will be the estimated inflation shock.

5.3.2 Prior Specification

We denote $\mathbf{Y}$ as the $T \times 4$ matrix of dependent variables, which vertically stacks the row vectors $\mathbf{Y}_t'$ (of dimension 1×4). Similarly, we denote $\mathbf{X}$ as the $T \times 13$ matrix of explanatory variables, which vertically stacks the row vectors $\mathbf{X}_t'$ (of dimension 1×13), containing lags of $\mathbf{Y}$ and deterministic terms. The deterministic components $\mathbf{X}_{\text{det}}$ and endogenous variables $\mathbf{X}_{\text{endo}}$ can be partitioned as:

$$\mathbf{X} = [\mathbf{X}_{\text{det}}, \mathbf{X}_{\text{endo}}]. \tag{20}$$

In a model with only a deterministic constant, $\mathbf{X}_{\text{det}}$ is a $T \times 1$ vector and $\mathbf{X}_{\text{endo}}$ is a $T \times 12$ matrix representing the endogenous variables. The reduced-form and identification parameters associated with the OLS estimate of model are given by:

$$\gamma = (\mathbf{X}_{\text{det}}'\mathbf{X}_{\text{det}})^{-1}\mathbf{X}_{\text{det}}'\mathbf{Y}\,, \tag{21}$$

$$\boldsymbol{\beta} = (\mathbf{X}_{\text{endo}}'\mathbf{X}_{\text{endo}})^{-1}\mathbf{X}_{\text{endo}}'\mathbf{Y}\,, \tag{22}$$

$$\mathbf{U} = \mathbf{Y} - \mathbf{X}_{\text{det}}\gamma - \mathbf{X}_{\text{endo}}\boldsymbol{\beta}\,, \tag{23}$$

$$\boldsymbol{\Sigma} = \frac{\mathbf{U}'\mathbf{U}}{T-k+1}\,, \tag{24}$$

$$\mathbf{Y} = (\mathbf{U}'\mathbf{U})^{-1}\mathbf{U}'\mathbf{M}\,, \tag{25}$$

$$\mathbf{V} = \mathbf{M} - \mathbf{U}\mathrm{Y}\,, \tag{26}$$

$$\boldsymbol{\Sigma}_V = \frac{\mathbf{V}'\mathbf{V}}{T-k}\,, \tag{27}$$

where Equation (21) are estimates for the coefficients associated with the deterministic terms, corresponding to the previously defined $\mathbf{c}'$, Equation (22) are the estimates for the slope coefficients corresponding to the endogenous variables, corresponding to the previously defined $[\boldsymbol{\Phi}_1', \boldsymbol{\Phi}_2', \boldsymbol{\Phi}_3']'$, Equation (23) are the residuals from the reduced-form VAR model, Equation (24) defines the variance-covariance matrix of the residuals, Equation (25) are the identification parameters obtained from the set of external (granular) instruments represented by the $T \times 3$ matrix $\mathbf{M}$, and Equation (26) states its corresponding identification errors (i.e., the residual). Finally, Equation (27) is the 3×3 variance-covariance matrix associated with the identification errors, where 4 is the number of endogenous variables in the VAR.

We assume that prior distributions of the slope parameters $\{\gamma, \boldsymbol{\beta}, \mathbf{Y}\}$ are Gaussian and those of the variance-covariance matrices $\{\boldsymbol{\Sigma}, \boldsymbol{\Sigma}_V\}$ to be an inverse-Wishart with degrees of freedom equal to $\bar{\nu} = 7$. In our implementation of the Gibbs sampler, we initialise the priors at their OLS estimates.

5.3.3 Posterior distributions, Model Estimation and Sampling

To economize on notation, we introduce set notation by defining the parameter space of the model Θ, the set of data used in the VAR model data Υ and the external instruments as M:

$$
\begin{aligned}
\Theta &= \{\boldsymbol{\beta}, \gamma, \boldsymbol{\Sigma}, \mathbf{Y}, \boldsymbol{\Sigma}_v\}\,, && (28)\\
\Upsilon &= \{\mathbf{Y}, \mathbf{X}\}\,, && (29)\\
M &= \{\mathbf{M}\}\,. && (30)
\end{aligned}
$$

Using Bayes rule, the joint likelihood function is given by:

$$
p(M, \Upsilon \mid \Theta) = p(M \mid \Upsilon, \Theta)p(\Upsilon \mid \Theta)\,, \tag{31}
$$

The posterior distribution of parameters is proportional to the product of the likelihood and the prior:

$$
p(\Theta \mid M, \Upsilon) \propto p(M, \Upsilon \mid \Theta)p(\Theta)\,. \tag{32}
$$

The VAR data and the granular instruments are jointly Gaussian, given by:

$$
\begin{pmatrix} vec(\mathbf{Y}) \\ vec(\mathbf{M}) \end{pmatrix} \Bigg| \Theta \sim \mathcal{N}\left(\begin{pmatrix} (\mathbf{I}_T \otimes \mathbf{X}_{endo})\boldsymbol{\beta} + (\mathbf{I}_T \otimes \mathbf{X}_{det})\gamma \\ \mathbf{0} \end{pmatrix}, \begin{bmatrix} \boldsymbol{\Phi}_{11} & \boldsymbol{\Phi}_{12} \\ \boldsymbol{\Phi}_{21} & \boldsymbol{\Phi}_{22} \end{bmatrix} \right), \tag{33}
$$

where $vec(\cdot)$ represents the vectorisation of a matrix, with dimensions of $vec(\mathbf{Y})$ and $vec(\mathbf{I})$ given by $4T \times 1$ and $3T \times 1$ respectively, and $\mathbf{I}_T$ represents the $T \times T$ identity matrix. The covariance block matrices are defined as:

$$
\begin{aligned}
\boldsymbol{\Phi}_{11} &= \boldsymbol{\Sigma} \otimes \mathbf{I}_T\,, && (34)\\
\boldsymbol{\Phi}_{22} &= \mathbf{Y}'\boldsymbol{\Sigma}\mathbf{Y} \otimes \mathbf{I}_T + \boldsymbol{\Sigma}_V \otimes \mathbf{I}_T\,, && (35)\\
\boldsymbol{\Phi}_{12} &= \boldsymbol{\Phi}_{21} = \boldsymbol{\Sigma}'\mathbf{Y} \otimes \mathbf{I}_T\,. && (36)
\end{aligned}
$$

Having defined these objects, the posterior distribution of the VAR slope coefficients $\boldsymbol{\beta}$ is given by:

$$
p(\boldsymbol{\beta} \mid \Upsilon, M, \Theta \setminus \boldsymbol{\beta}) \propto \mathcal{N}(\mathbf{D}^{-1}\mathbf{d}, \mathbf{D}^{-1}), \tag{37}
$$

where the posterior mean and variances of $\boldsymbol{\beta}$ are defined, respectively, as:

$$
\begin{aligned}
\mathbf{D} &= (\mathbf{I}_T \otimes \mathbf{X}_{endo})'\left(\boldsymbol{\Phi}_{11} - \boldsymbol{\Phi}_{12}\boldsymbol{\Phi}_{22}^{-1}\boldsymbol{\Phi}_{21}\right)^{-1}(\mathbf{I}_T \otimes \mathbf{X}_{endo})\,, && (38)\\
\mathbf{d} &= (\mathbf{I}_T \otimes \mathbf{X}_{endo})'\left(\boldsymbol{\Phi}_{11} - \boldsymbol{\Phi}_{12}\boldsymbol{\Phi}_{22}^{-1}\boldsymbol{\Phi}_{21}\right)^{-1}\left[vec(\mathbf{Y}) - (\mathbf{I}_T \otimes \mathbf{X}_{det})\gamma - \boldsymbol{\Phi}_{12}\boldsymbol{\Phi}_{22}^{-1}vec(\mathbf{M})\right]\,. && (39)
\end{aligned}
$$

Taking the covariance block matrices as given from Equation (34) to Equation (36), and taking the parameter γ as given all from the previous draw from the Gibbs sampler, Equations (38) and (39) provide an update to the posterior distribution of $\boldsymbol{\beta}$ given the data $\{Y, M\}$, for which the values of $\boldsymbol{\beta}$ can be updated by drawing from the posterior distribution in Equation (37).

Turning now to the deterministic coefficients γ, the posterior distribution is expressed as:

$$p(\gamma \mid Y, M, \Theta \setminus \boldsymbol{\beta}) \sim \mathcal{N}(\mathbf{F}^{-1}\mathbf{f}, \mathbf{F}^{-1}), \tag{40}$$

where the posterior mean and variances of γ are defined as:

$$\mathbf{F} = (\mathbf{I}_T \otimes \mathbf{X}_{det})' \left(\boldsymbol{\Phi}_{11} - \boldsymbol{\Phi}_{12}\boldsymbol{\Phi}_{22}^{-1}\boldsymbol{\Phi}_{21}\right)^{-1} (\mathbf{I}_T \otimes \mathbf{X}_{det}), \tag{41}$$

$$\mathbf{f} = (\mathbf{I}_T \otimes \mathbf{X}_{det})' \left(\boldsymbol{\Phi}_{11} - \boldsymbol{\Phi}_{12}\boldsymbol{\Phi}_{22}^{-1}\boldsymbol{\Phi}_{21}\right)^{-1} \left[vec(\mathbf{Y}) - (\mathbf{I}_T \otimes \mathbf{X}_{endo})\boldsymbol{\beta} - \boldsymbol{\Phi}_{12}\boldsymbol{\Phi}_{22}^{-1} vec(\mathbf{M})\right]. \tag{42}$$

Thus, taking the covariance block matrices as given from Equation (34) to Equation (36), and taking the parameter $\boldsymbol{\beta}$ as given from the previous draw from the Gibbs sampler, Equations (41) and (42) provide an update to the posterior distribution of γ given the data $\{Y, M\}$, for which the values of γ can be updated by drawing from the posterior distribution in Equation (40).

For the covariance matrix $\boldsymbol{\Sigma}$, its posterior distribution is given by:

$$p(\boldsymbol{\Sigma} \mid Y, \Theta \setminus \boldsymbol{\Sigma}) \sim \mathcal{IW}(\mathbf{U}'\mathbf{U}, T + \bar{v}) \tag{43}$$

where $\mathbf{U} = \mathbf{Y} - \mathbf{X}_{\text{endo}}\boldsymbol{\beta} - \mathbf{X}_{\text{det}}\gamma$ and $\bar{v} = 7$. Thus, the matrix of residuals given the updated draws of $\boldsymbol{\beta}$ and γ are used to update the posterior distribution of $\boldsymbol{\Sigma}$, for which an update to the value of $\boldsymbol{\Sigma}$ is obtained by drawing from the distribution in Equation (43).

Next, we turn to the posterior distribution of the identification parameters $\mathbf{Y}$. This is given by:

$$p(\mathbf{Y} \mid Y, M, \Theta \setminus \mathbf{Y}) \sim \mathcal{N}(\mathbf{J}^{-1}\mathbf{j}, \mathbf{J}^{-1}), \tag{44}$$

where the posterior mean and variances of $\mathbf{Y}$ are defined as:

$$\mathbf{J} = \boldsymbol{\Sigma}_V^{-1} \otimes \mathbf{U}'\mathbf{U} \tag{45}$$

$$\mathbf{j} = (\boldsymbol{\Sigma}_V^{-1} \otimes \mathbf{U}') \operatorname{vec}(\mathbf{M}) \tag{46}$$

Hence, given the update of $\mathbf{U}$ and the variance-covariance of the identification errors $\boldsymbol{\Sigma}_V$, the posterior distribution of $\mathbf{Y}$ is updated by the data M. An update to $\mathbf{Y}$ can thus be obtained by drawing from the posterior distribution in Equation (44).

Finally, the posterior of the instrument variance-covariance matrix $\boldsymbol{\Sigma}_V$ is given by:

$$p(\boldsymbol{\Sigma}_V \mid M, \Theta) \sim \mathcal{IW}(\mathbf{V}'\mathbf{V}, T), \tag{47}$$

where $\mathbf{V}$ is the instrument residual matrix $\mathbf{V} = \mathbf{M} - \mathbf{U}\mathbf{Y}$. The update of $\mathbf{Y}$ from Equation (44) implies that the posterior distribution of $\boldsymbol{\Sigma}_V$ can be updated in the presence of data $\mathbf{U}$ and the granular instruments **M**.

For the baseline specification, the posterior distribution is simulated using 1,500,000 draws from the MCMC sampler; the first 500,000 are discarded as burn-in, and the remaining chain is thinned by a factor of 100, leaving 10,000 draws for inference. Results are presented as the median of the 10,000 retained draws, and credible intervals are computed using standard Bayesian Monte Carlo methods. This process ensures that the convergence diagnostics on the parameters in the set Θ are passed. For robustness checks or extensions, a shorter chain of 500,000 draws is used, where the first 400,000 are discarded as burn-in, and the remaining draws are thinned by a factor of 10, leaving 10,000 draws for inference.

5.4 Bayesian Structural VAR with Sign and Zero Restrictions

In this section, we turn to our last identification strategy that implements a Bayesian structural VAR with sign and zero restrictions. We follow the algorithm of Arias et al. (2018) to independently draw from a family of conjugate posterior distributions over the structural parameterisation conditional on the sign and zero restrictions. In particular, the algorithm that we implement independently draws from the normal-inverse-Wishart distribution over the orthogonal reduced-form, and accepting the draws only when the sign and zero restrictions hold. We only provide a sketch of the main steps, emphasising the restrictions we impose and numerical details from the sampler, and invite readers to follow Arias et al. (2018) for further details about the algorithm.

We consider a structural vector autoregression model represented by:

$$\mathbf{Y}_t'\mathbf{A}_0 = \mathbf{X}_t'\mathbf{A}_+ + \boldsymbol{\varepsilon}_t', \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(0, \mathbf{I}_n), \tag{48}$$

where $\boldsymbol{\varepsilon}_t$ is an 4×1 vector of latent structural shocks, and $\mathbf{A}_0$ (of dimensions 4×4) and $\mathbf{A}_+$ (of dimensions 13×4) are the structural-form parameters that identify these structural shocks. We assume the condition that $\mathbf{A}_0$ is invertible. Conditional on the data $\mathbf{X}_t$, the structural shocks are also assumed to be Gaussian with mean zero and covariance matrix $\mathbf{I}_n$, the $n \times n$ identity matrix. The orthogonal reduced-form representation of the VAR is therefore given by:

$$\mathbf{Y}_t' = \mathbf{X}_t'\mathbf{B} + \mathbf{u}_t', \quad \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{49}$$

where $\mathbf{B} = [\mathbf{c}', \boldsymbol{\Phi}_1', \boldsymbol{\Phi}_2', \boldsymbol{\Phi}_3']' = \mathbf{A}_+\mathbf{A}_0^{-1}$, $\mathbf{u}_t' = \boldsymbol{\varepsilon}_t'\mathbf{A}_0^{-1}$, and $\mathbb{E}[\mathbf{u}_t\mathbf{u}_t'] = \boldsymbol{\Sigma} = (\mathbf{A}_0\mathbf{A}_0')^{-1}$. The matrices $\{\mathbf{B}, \boldsymbol{\Sigma}\}$ are the reduced-form parameters, while $\{\mathbf{A}_0, \mathbf{A}_+\}$ are the structural-form parameters.

5.4.1 Prior and Posterior Distributions

For Bayesian inference we adopt a normal–inverse Wishart prior. In particular, we assume that the prior distribution of $\boldsymbol{\Sigma}$ follows:

$$\boldsymbol{\Sigma} \sim \mathcal{IW}(\bar{\boldsymbol{\Phi}}, \bar{v}), \tag{50}$$

and, conditional on $\boldsymbol{\Sigma}$, the prior distribution for $\mathbf{B}$ is specified as:

$$\text{vec}(\mathbf{B}) \mid \boldsymbol{\Sigma} \sim \mathcal{N}\Big(vec(\bar{\boldsymbol{\Psi}}),\, \boldsymbol{\Sigma} \otimes \bar{\boldsymbol{\Omega}}\Big). \tag{51}$$

We impose diffuse priors by setting $\bar{\nu} = 0$, $\bar{\boldsymbol{\Omega}}^{-1} = \mathbf{0}_{13\times13}$, $\bar{\boldsymbol{\Psi}} = \mathbf{0}_{13\times4}$, $\bar{\boldsymbol{\Phi}} = \mathbf{0}_{4\times4}$, where m and n conforms to the dimensions of the VAR, and let the data inform the posterior distribution entirely. Given VAR data $\{\mathbf{Y}, \mathbf{X}\}$, where $\mathbf{Y} = [\mathbf{Y}_1, ..., \mathbf{Y}_T]'$ and $\mathbf{X} = [\mathbf{X}_1, ..., \mathbf{X}_T]'$ and the priors from Equations (50) to Equation (51), the posterior update to the normal-inverse Wishart distribution follows:

$$\tilde{\nu} = T + \bar{\nu}, \tag{52}$$

$$\tilde{\boldsymbol{\Omega}} = \left(\mathbf{X}'\mathbf{X} + \bar{\boldsymbol{\Omega}}^{-1}\right)^{-1}, \tag{53}$$

$$\tilde{\boldsymbol{\Psi}} = \tilde{\boldsymbol{\Omega}}\left(\mathbf{X}'\mathbf{Y} + \bar{\boldsymbol{\Omega}}^{-1}\bar{\boldsymbol{\Psi}}\right), \tag{54}$$

$$\tilde{\boldsymbol{\Phi}} = \mathbf{Y}'\mathbf{Y} + \bar{\boldsymbol{\Phi}} + \bar{\boldsymbol{\Psi}}'\,\bar{\boldsymbol{\Omega}}^{-1}\,\bar{\boldsymbol{\Psi}} - \tilde{\boldsymbol{\Psi}}'\tilde{\boldsymbol{\Omega}}^{-1}\tilde{\boldsymbol{\Psi}}, \tag{55}$$

after which we draw the reduced-form parameters $\{\mathbf{B}, \boldsymbol{\Sigma}\}$ independently from the posterior distributions:

$$\boldsymbol{\Sigma} \mid \mathbf{Y}, \mathbf{X} \sim \mathcal{IW}(\tilde{\boldsymbol{\Phi}}, \tilde{\nu}), \tag{56}$$

$$vec(\mathbf{B}) \mid \boldsymbol{\Sigma}, \mathbf{Y}, \mathbf{X} \sim \mathcal{N}\left(vec(\tilde{\boldsymbol{\Psi}}),\, \boldsymbol{\Sigma} \otimes \tilde{\boldsymbol{\Omega}}\right). \tag{57}$$

5.4.2 Identification via Sign and Zero Restrictions

The structural identification is achieved by imposing restrictions on the contemporaneous impulse responses derived from the structural representation in Equation (48). We follow Theorem 4 of Arias et al. (2018) in drawing random, orthogonal matrices $\mathbf{Q}$ of dimension 4×4 from the uniform distribution over $\mathcal{O}(4)$, and given a Cholesky decomposition of $\boldsymbol{\Sigma}$ such that $\boldsymbol{\Sigma} = \mathbf{P}\,\mathbf{P}'$, we can obtain a rotation of this decomposition such that it is a candidate draw of the impact response matrix of the structural shocks:

$$\mathbf{A}_0^{-1} = \mathbf{P}\,\mathbf{Q}, \tag{58}$$

and the corresponding impulse response functions at horizon h are given by:

$$\mathbf{L}_h = \mathbf{A}_0^{-1}\boldsymbol{\Phi}_h = \mathbf{P}\,\mathbf{Q}\,\boldsymbol{\Phi}_h. \tag{59}$$

In our application, we impose the sign and zero restrictions on the contemporaneous impulse response functions $\mathbf{L}_0$, restricting the effects of the structural shocks on impact. These restrictions are the following:

Hedge fund demand shock: we impose the sign restriction that a positive demand shock from hedge funds in the short-maturity swap market *increase* both the net purchases of swaps traded against dealer banks in this market and the short-term inflation swap rate, whilst having zero effects on the corresponding net purchases of swaps and its associated swap rate in the long-maturity swap market. These restrictions are represented by the following sign (**S**) and zero (**Z**) restriction matrices:

$$\mathbf{S}_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}, \quad \mathbf{Z}_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{60}$$

Pension fund demand shock: we impose the sign restriction that a positive demand shock from pension funds in the long-maturity swap market *increase* both the net purchases of swaps traded against dealer banks in this market and the long-term inflation swap rate, whilst having zero effects on the corresponding net purchases of swaps and its associated swap rate in the short-maturity swap market. These restrictions are represented by the following sign and zero restriction matrices:

$$\mathbf{S}_2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \quad \mathbf{Z}_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \tag{61}$$

Dealer bank supply shock: we impose the sign restriction that a supply shock from dealer banks *decreases* the net purchases of swaps traded against hedge funds in the short-maturity swap market and those traded against pension funds in the long-maturity swap market, leading to an *increase* in inflation swap rates in both markets. That is, we rely on sign restrictions entirely to identify the supply shock and do not impose zero restrictions.

$$\mathbf{S}_3 = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{62}$$

Inflation shock: we impose the sign restriction that a inflation shock resembles a supply shock in the long-maturity swap market and a demand shock in the short-maturity swap market. Thus, this imposes sign restrictions such that the shock *increases* both the net purchases and associated inflation swap rate in the short-maturity market, whilst *decreasing* net purchases in the long-maturity swap market. The sign restrictions we impose are (note that the last restriction on long-maturity swap rate is redundant):

$$\mathbf{S}_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \end{pmatrix}. \tag{63}$$

5.4.3 Implementation of Restrictions and Inference

We proceed with estimation by drawing the reduced-form parameters from the joint posterior $p(\mathbf{B}, \boldsymbol{\Sigma} \mid \mathbf{Y}, \mathbf{X})$ as described in Section 5.4.1, and keeping the draws only if their transformed structural-form parameterisations satisfy the sign and zero restrictions that we impose in Section 5.4.2. For each draw i, the reduced-form covariance matrix $\boldsymbol{\Sigma}^{(i)}$ (a 4×4 matrix) is decomposed as

$$\boldsymbol{\Sigma}^{(i)} = \mathbf{P}^{(i)} \left(\mathbf{P}^{(i)}\right)', \tag{64}$$

with $\mathbf{P}^{(i)} \in \mathbb{R}^{4\times 4}$. An auxiliary variable $\mathbf{w}$ is constructed such that its elements are drawn from the uniform distribution on the unit sphere and are subsequently used to construct an orthogonal matrix:

$$\mathbf{Q}^{(i)} \in \mathbb{R}^{4\times 4}, \quad \mathbf{Q}^{(i)}\left(\mathbf{Q}^{(i)}\right)' = \mathbf{I}_4. \tag{65}$$

The transformation from the reduced-form parameters (without any restrictions imposed) to the structural parameters is achieved via the mapping:

$$\boldsymbol{\theta}^{(i)} = \begin{pmatrix} \mathrm{vec}\left(\mathbf{A}_0^{(i)}\right) \\ \mathrm{vec}\left(\mathbf{A}_+^{(i)}\right) \end{pmatrix} = f_h^{-1}\Big(vec\left(\mathbf{B}^{(i)}\right), vec\left(\boldsymbol{\Sigma}^{(i)}\right), \mathbf{w}\Big). \tag{66}$$

This mapping constructs the structural matrices $\mathbf{A}_0$ and $\mathbf{A}_+$ so that

$$\mathbf{B}^{(i)} = \left(\mathbf{A}_0^{(i)}\right)^{-1} \mathbf{A}_+^{(i)} \quad \text{and} \quad \boldsymbol{\Sigma}^{(i)} = \left(\mathbf{A}_0^{(i)}\right)^{-1} \left(\left(\mathbf{A}_0^{(i)}\right)^{-1}\right)'. \tag{67}$$

In particular, we have

$$\left(\mathbf{A}_0^{(i)}\right)^{-1} = \mathbf{P}^{(i)}\,\mathbf{Q}^{(i)}, \tag{68}$$

and given the impulse response multipliers $\{\boldsymbol{\Phi}_h\}_{h\geq 0}$, the impulse responses are computed as

$$\mathbf{L}_h^{(i)} = \left(\mathbf{A}_0^{(i)}\right)^{-1} \boldsymbol{\Phi}_h, \tag{69}$$

with the contemporaneous impulse response given by $\mathbf{L}_0^{(i)} = \left(\mathbf{A}_0^{(i)}\right)^{-1}$ since $\boldsymbol{\Phi}_0 = \mathbf{I}_4$. The sign and zero restrictions on $\mathbf{L}_0^{(i)}$ are then imposed by requiring that the following conditions hold for the accepted draws:

For hedge fund demand shock:

$$\mathbf{S}_1\,\mathbf{l}_1^{(i)} = \begin{pmatrix} \ell_{11}^{(i)} \\ \ell_{21}^{(i)} \end{pmatrix} \geq 0, \quad \text{and} \quad \mathbf{Z}_1\,\mathbf{l}_1^{(i)} = \begin{pmatrix} \ell_{31}^{(i)} \\ \ell_{41}^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{70}$$

For pension fund demand shock:

$$\mathbf{S}_2\,\mathbf{l}_2^{(i)} = \begin{pmatrix} \ell_{32}^{(i)} \\ \ell_{42}^{(i)} \end{pmatrix} \geq 0, \quad \text{and} \quad \mathbf{Z}_2\,\mathbf{l}_2^{(i)} = \begin{pmatrix} \ell_{12}^{(i)} \\ \ell_{22}^{(i)} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \tag{71}$$

For dealer supply shock:

$$\mathbf{S}_3\,\mathbf{l}_3^{(i)} = \begin{pmatrix} -\ell_{13}^{(i)} \\ \ell_{23}^{(i)} \\ -\ell_{33}^{(i)} \\ \ell_{43}^{(i)} \end{pmatrix} \geq 0. \tag{72}$$

For the inflation shock:

$$\mathbf{S}_4\,\mathbf{l}_4^{(i)} = \begin{pmatrix} \ell_{14}^{(i)} \\ \ell_{24}^{(i)} \\ -\ell_{34}^{(i)} \end{pmatrix} \geq 0. \tag{73}$$

In these, $\mathbf{l}_1^{(i)}, \mathbf{l}_2^{(i)}, \mathbf{l}_3^{(i)}$, and $\mathbf{l}_4^{(i)}$ denote the first, second, third, and fourth columns of $\mathbf{L}_0^{(i)}$, respectively.

A draw i is accepted if and only if *all* imposed sign and zero restrictions are satisfied. For each accepted draw, an importance sampling weight is computed to adjust for both the transformation from unconstrained reduced-form parameters to the structural parameters *and* the reduction in the admissible parameter space due to zero restrictions, such that inference is not dominated by only a few draws. Specifically, the mapping

$$f_h^{-1} : \; \Big(\operatorname{vec}\big(\mathbf{B}^{(i)}\big), \operatorname{vec}\big(\boldsymbol{\Sigma}^{(i)}\big), \mathbf{w}\Big) \;\mapsto\; \boldsymbol{\theta}^{(i)} \tag{74}$$

transforms the unconstrained parameters into the structural parameter vector $\boldsymbol{\theta}^{(i)}$, which encapsulates both the contemporaneous structural impact matrix $\mathbf{A}_0$ and the lag coefficient matrices $\mathbf{A}_+$. As this mapping is one-to-one, it introduces a Jacobian factor $f_h\big(\boldsymbol{\theta}^{(i)}\big)$ that corrects the density during the transformation. Furthermore, the imposition of zero restrictions reduces the effective volume of the parameter space. This reduction is quantified by a volume element:

$$g\Big(f_h\big(\boldsymbol{\theta}^{(i)}\big) \mid \mathbf{Z}\Big), \tag{75}$$

which depends explicitly on the zero restriction matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$. Consequently, the importance sampling weight for the i^{th} accepted draw is given by

$$\omega^{(i)} \;\propto\; \exp\Big\{\log f_h\big(\boldsymbol{\theta}^{(i)}\big) \;-\; \log g\Big(f_h\big(\boldsymbol{\theta}^{(i)}\big) \mid \mathbf{Z}\Big)\Big\}. \tag{76}$$

The effective sample size is thus given by:

$$\frac{\left(\sum_{i=1}^{N^*} \omega^{(i)}\right)^2}{\sum_{i=1}^{N^*} (\omega^{(i)})^2}, \tag{77}$$

where N^* is the total number of draws that satisfies the sign and zero restrictions.

For inference, we implement 10 million iterations from the importance sampler (corresponding to Algorithm 3 in Arias et al. (2018)), of which a total of 35,160 draws satisfy the sign and zero restrictions imposed. Given the importance sampling weights associated with each accepted draw, the effective sample size according to Equation (77) is 9,955. This corresponds to an effective sample size as a share of total draws satisfying the sign and zero restrictions equal to 0.28. Our baseline results for inference use the median of these 9,955 effective draws from the importance sampler along with its 68% and 90% credible intervals.

6 Additional empirical results

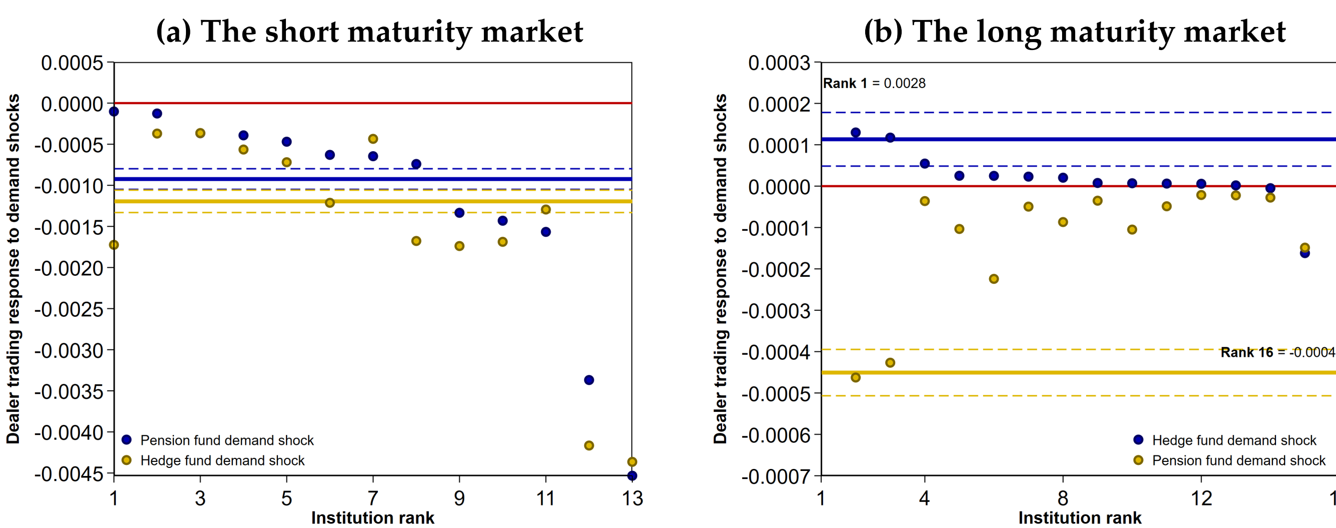
This section includes additional empirical results.

6.1 Testing the desk separation assumption directly

Under the desk separation assumption, a demand shock that emerges from one market should not have a bearing on dealers' supply of inflation swaps in another market, at least at a daily frequency. We test this assumption by running seemingly unrelated regressions of $q_{b,i,t}/a_{b,i,t}$ on $\varepsilon_{h,t}$ and $Q_{b,i,t}/q_{b,i,t}$ on $\varepsilon_{f,t}$, using the structural demand shocks $\varepsilon_{h,t}$ and $\varepsilon_{f,t}$ identified using a granular instrument.

Figure 9 shows that the pooled coefficient estimate is close to zero, closely aligned with the fact that individual estimates obtained from the trading activities of each dealer bank are also extremely small.

Figure 9 TEST OF THE DESK SEPARATION ASSUMPTION



NOTE: The thick blue and yellow lines are pooled coefficient estimates for dealers' trading responses against hedge fund and pension fund demand shocks respectively, where these frictional shocks are the median estimates obtained from the using the GIV identification strategy. The dashed lines are their 95% credible sets where standard errors clustered at the institutional-level (where each is identified by a legal entity identifier). Sample goes from January 2019 to February 2023. SOURCE: DTCC Trade Repository and authors' calculations.

6.2 Movements around large shocks

Figure 10 presents the complete set of forecast error decompositions for the two episodes discussed in the main text with large changes in expected inflation for all four series in the VAR. These decompositions are the cumulative contributions of the inflation shock and demand and supply shocks (i.e., aggregating across dealer supply shocks, pension fund and hedge fund demand shocks) to the forecast error of both prices and quantities in each market, given an initial starting date. The left column decomposes each series' forecast error from just before the start of the Covid pandemic from February 2020, while the right column provides this decomposition since the start of the Ukraine War from February 2022.

Turning to the LDI episode, figure 11 repeats the comparison with bid0ask spreads for this episode and shows that the pattern of dealer supply shocks we see closely matches the evolution of bid-ask spreads in the swap market. These bid-ask spreads in the swap market also only started widening after the immediate phase of the selling had become apparent and issues with counterparty risk became salient.

6.3 Adding MPC dates

In our main analysis our heteroskedasticity-based identification strategy relied on inflation data releases and a policy announcement that influenced energy prices. This left us with 51 events in total. Monetary policy announcements constitute another set of dates that may convey the central bank's assessment of inflation, leading to first order changes in the variance of our data. Hence, in this section we extend the set of dates to include the 33 monetary policy announcements between January 2019 and February 2023. Figure 12 below presents a comparison to the impulse response to a inflation shock with and without the inclusion of monetary policy meeting dates. There are negligible differences.

6.4 Estimated impulse responses to frictional shocks

Figure 13 shows the estimated dynamic responses to the demand and supply shocks with sign restrictions. Given the sign restrictions, these conform with the standard responses one would expect from shocks to supply and demand.

6.5 Impulse response functions including insurance companies

We now report the estimated impulse responses when insurance companies quantities are added to the pension fund sector.

The responses to the demand and supply shocks under our first identification strategy — sign restrictions using the high-frequency data — are reported in Figure 14. Comparing these estimates to Figure 13, we see that the results from the main text are preserved even when adding the variation that originates from insurance companies. The responses to the inflation shocks are reported in Figure 15. Comparing them with the baseline results in the main text, again we find that the results are very similar whether insurance companies are included or not.

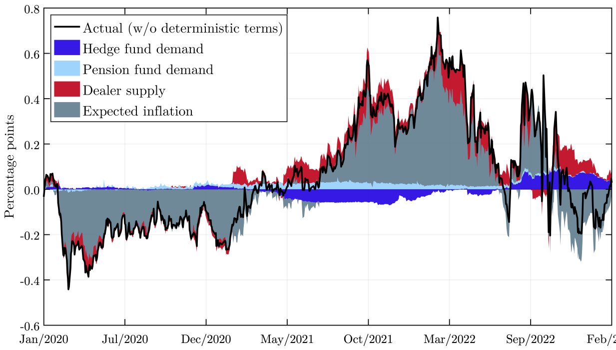
6.6 Responsiveness of trading positions to inflation shock in the long-maturity market

The paper showed how trade by dealers in the short-maturity market segments was sensitive to instittuion-level expected inflation. Using the same 3SLS regression of institutional-level trading activity. Figure 16 shows coefficients estimated for the activity of both dealers and pension funds on the long-maturity market segment. It shows three alternatives, corresponding to the thee identification strategies for the inflation shocks. The bold blue line represents the pooled coefficient estimate, along with the 95% confidence interval, obtained by clustering standard errors at the institutional level.

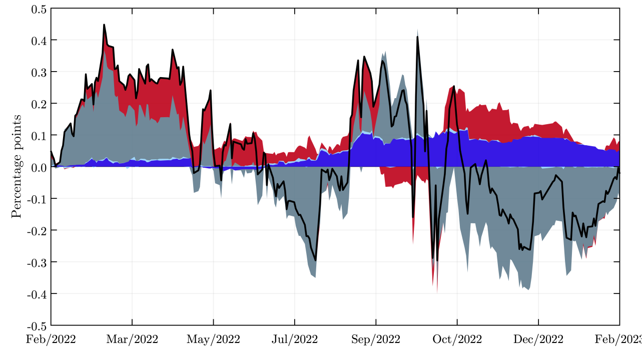
Figure 10 Full forecast error decompositions around the two episodes

Decomposition of long-maturity inflation swap breakeven rates

(a) Covid-19 period

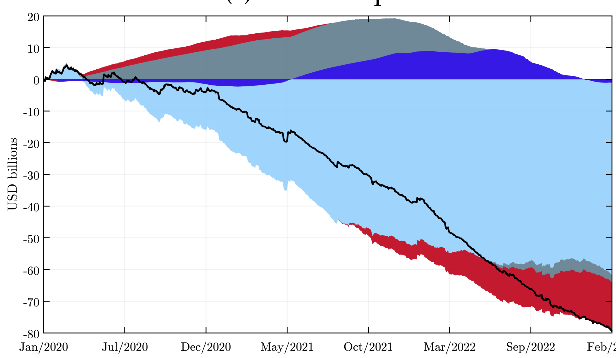


(b) Ukraine Invasion

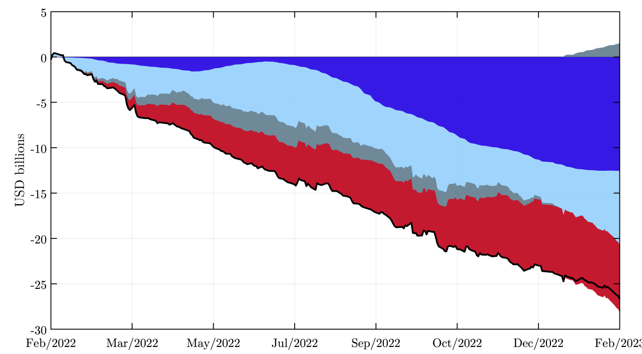


Decomposition of long-maturity net purchases (Cumulated)

(c) Covid-19 period

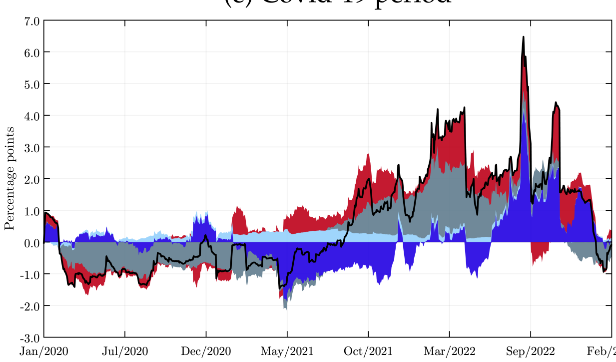


(d) Ukraine Invasion

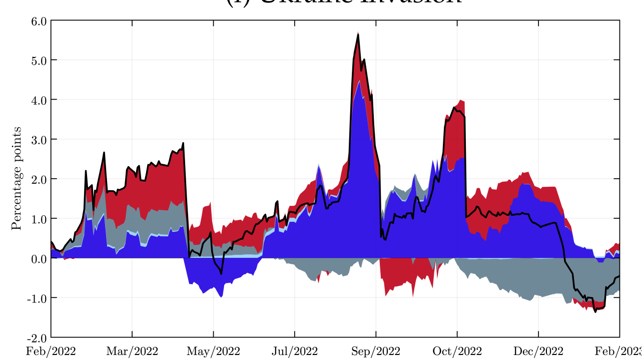


Decomposition of short-maturity inflation swap breakeven rates

(e) Covid-19 period

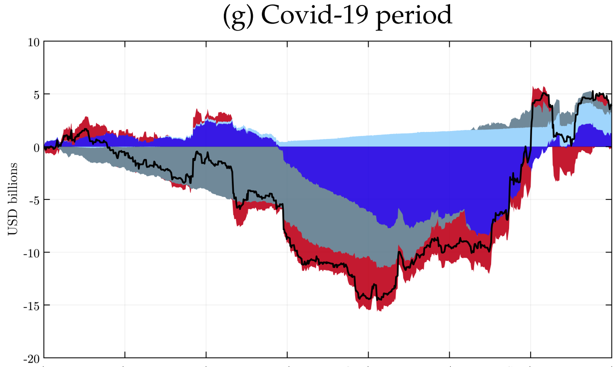


(f) Ukraine Invasion

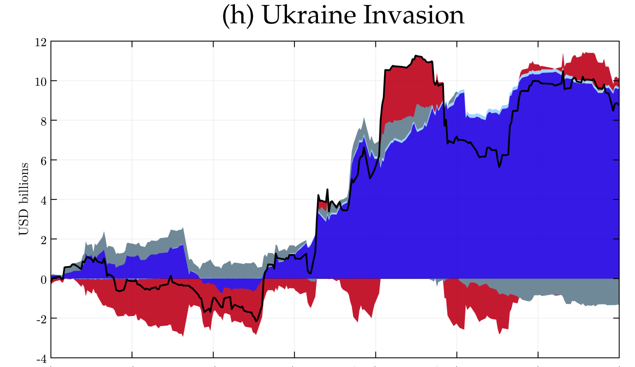


Decomposition of short-maturity net purchases (Cumulated)

(g) Covid-19 period

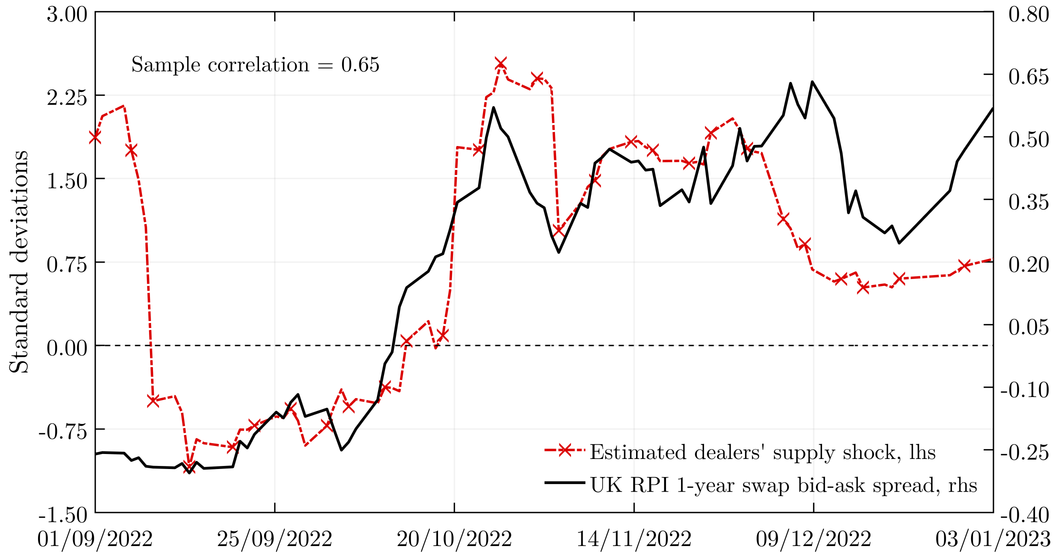


(h) Ukraine Invasion



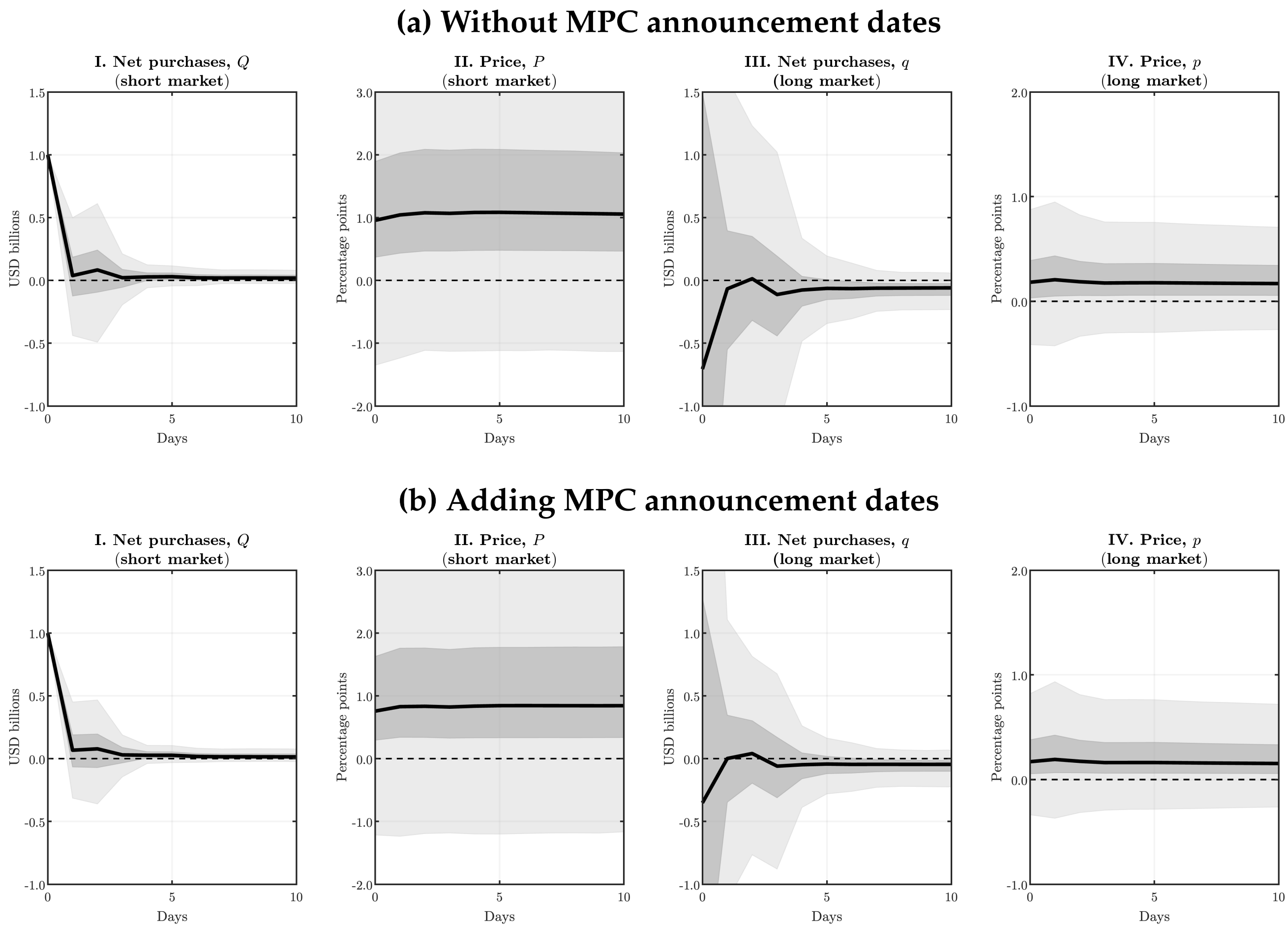
Note: This figure shows the forecast error decompositions of prices and quantities across both segments of the market for the two key episodes described in the main text: Covid-19 period (beginning early 2020) and the Ukraine invasion (early February 2022). Shaded area plots measure the %pp (USD billions) contribution to prices (quantities) by the median demand, supply and inflation shocks identified using the sign restrictions strategy. "Actual" refers to the associated price or quantity measure, without the estimated deterministic components, and therefore precisely equal to the vertical sum of the shaded plot areas.

Figure 11 COMPARISON BETWEEN DEALER SUPPLY SHOCKS AND MARKET BID-ASK SPREADS – AUTUMN 2022



NOTE: This figure compares the bid-ask spreads quoted on a 1-year zero-coupon UK RPI swap (right scale), and the cumulative contribution to the short-maturity inflation swap rate by the dealer supply shock estimated using the sign restrictions identification strategy (left scale). Dealers' supply shocks are obtained as a median of the draws from the importance sampler. Data on UK RPI swap bid-ask spreads are available from Bloomberg at a daily-frequency. SOURCE: Bloomberg and authors' calculations.

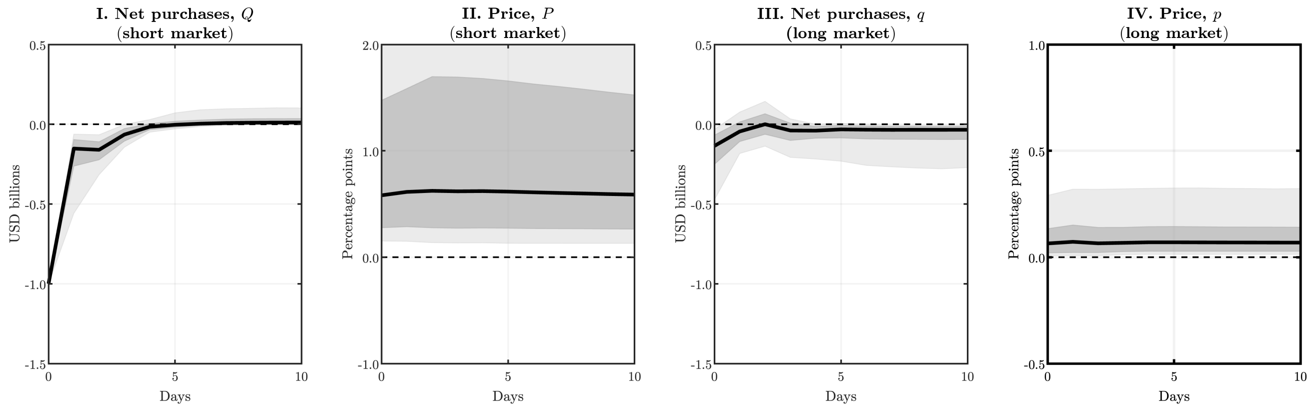
Figure 12 IMPULSE RESPONSE FUNCTIONS TO AN INFLATION SHOCK:
ESTIMATES FROM A HETEROSKEDASTICITY-BASED IDENTIFICATION STRATEGY



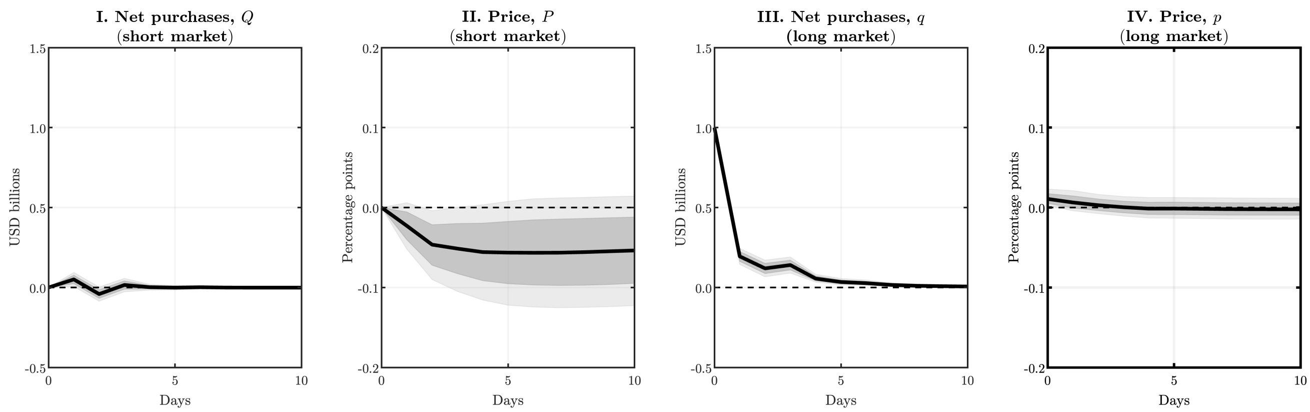
NOTE: Impulse response functions to a inflation shock scaled to raise quantities in the short market (i.e. purchases of inflation protection by hedge funds) by $1bn. Panel (a) presents the estimates with the set of 51 dates that coincide with the benchmark results from the main text in Figure 10. Panel (b) additionally includes the 33 monetary policy announcement dates i.e., a total of 84 dates to define the high covariance sample in the data. The bold line indicates the median of the draws from the sampler and shaded bands are its 68% and 90% credible intervals.

Figure 13 ESTIMATED IMPULSE RESPONSE FUNCTIONS TO DEMAND & SUPPLY SHOCKS

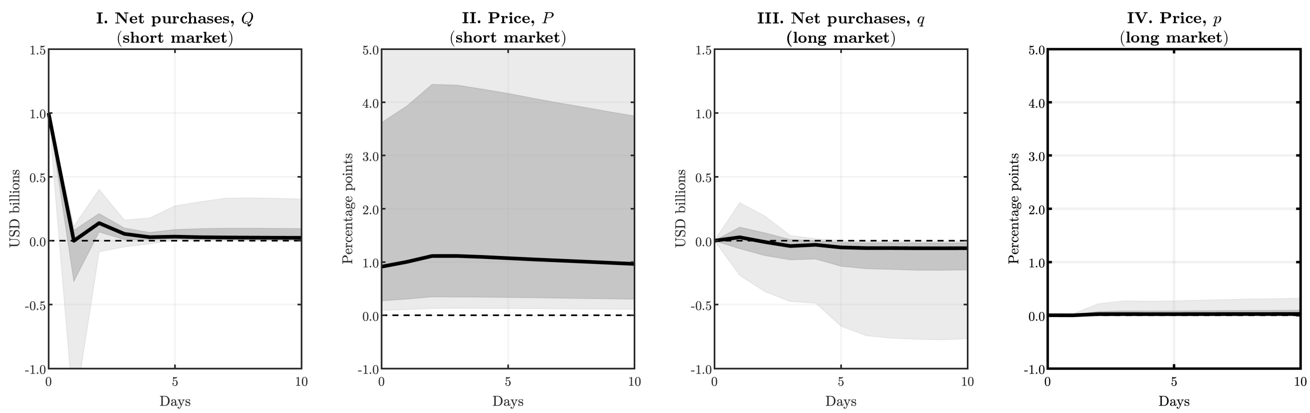
(a) Dealer supply shock ($\varepsilon_{b,t}$)



(b) Pension fund demand shock ($\varepsilon_{f,t}$)



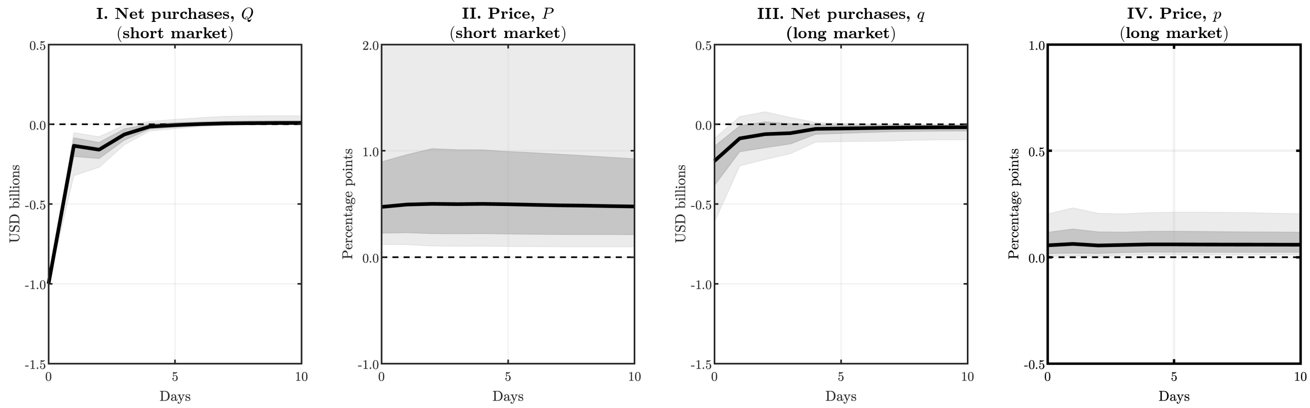
(c) Hedge fund demand shock ($\varepsilon_{h,t}$)



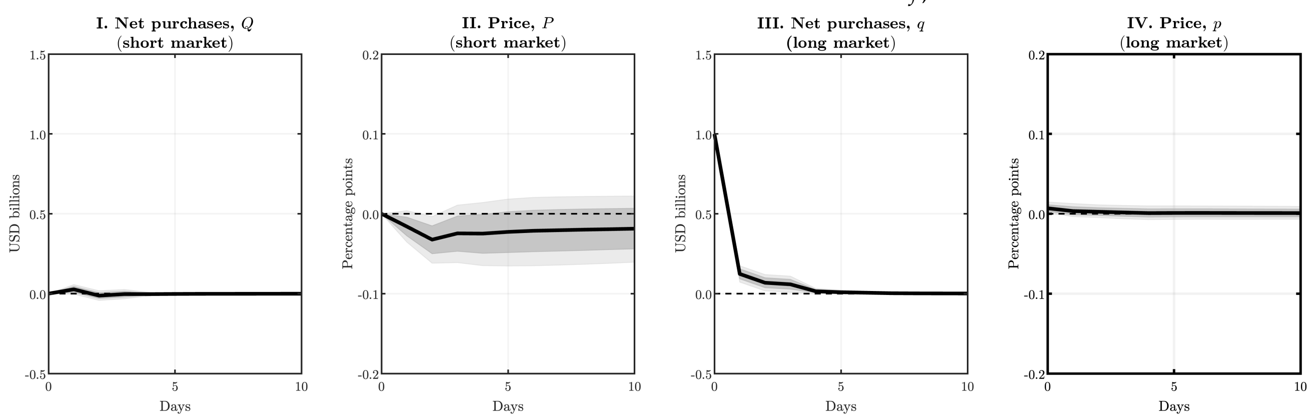
NOTE: Panel (a) presents impulse response functions to a dealer supply shock scaled to lower quantities in the short market (i.e. purchases of inflation protection by hedge funds) by \$1bn. Panel (b) presents impulse response functions to a pension fund demand shock scaled to raise quantities in the long market (i.e. purchases of inflation protection by pension funds) by \$1bn. Panel (c) presents impulse response functions to a hedge fund demand shock scaled to raise quantities in the short market by \$1bn. Across all panels, these demand and supply shocks are identified using the sign restrictions identification strategy. The posterior is simulated using 10 million draws using the sampler of Arias, Rubio-Ramírez, and Waggoner (2018) that selects 9,955 importance sampling draws for inference. The bold line indicates the median of the draws from the sampler and shaded bands are its 68% and 90% credible intervals.

Figure 14 Estimated impulse response functions to demand & supply shocks: Adding insurers in the long market

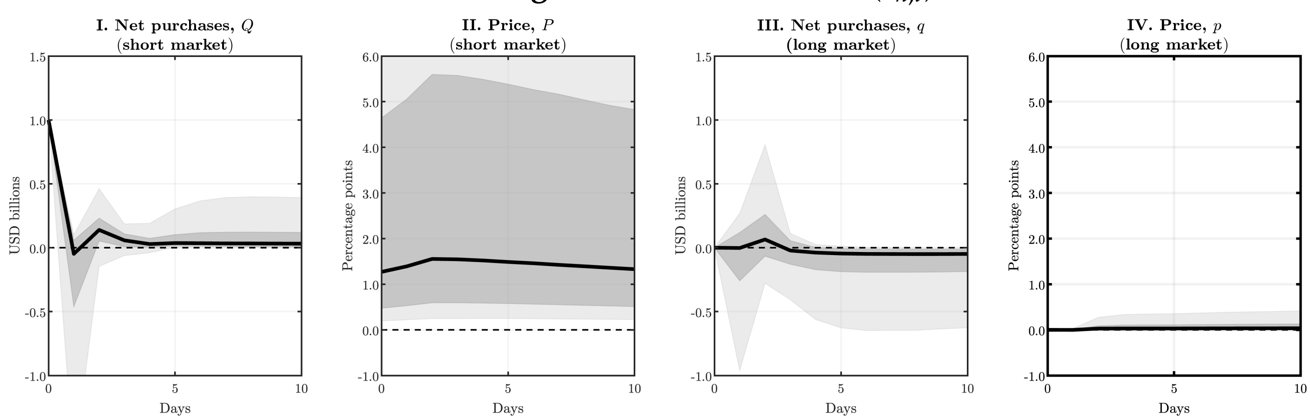
(a) Dealer supply shock ($\varepsilon_{b,t}$)



(b) Pension fund demand shock ($\varepsilon_{f,t}$)



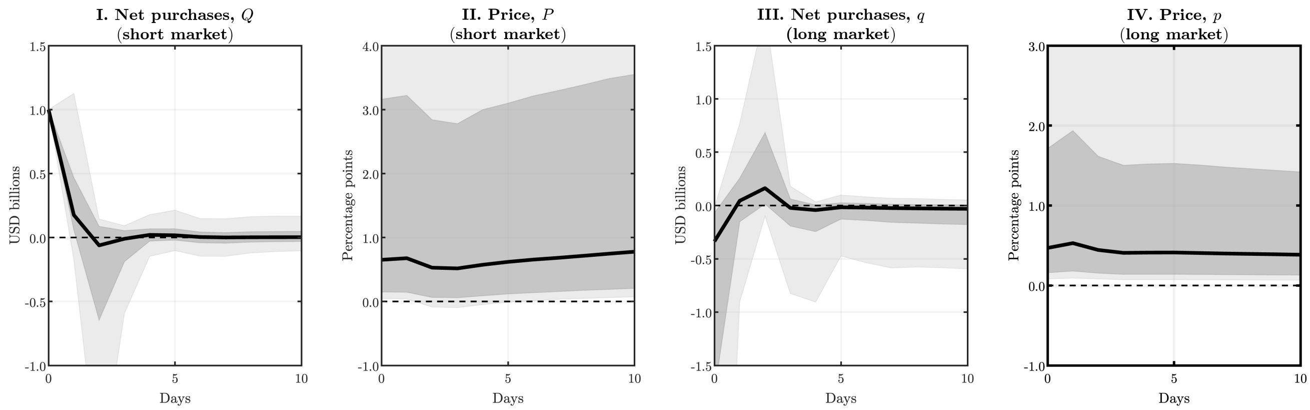
(c) Hedge fund demand shock ($\varepsilon_{h,t}$)



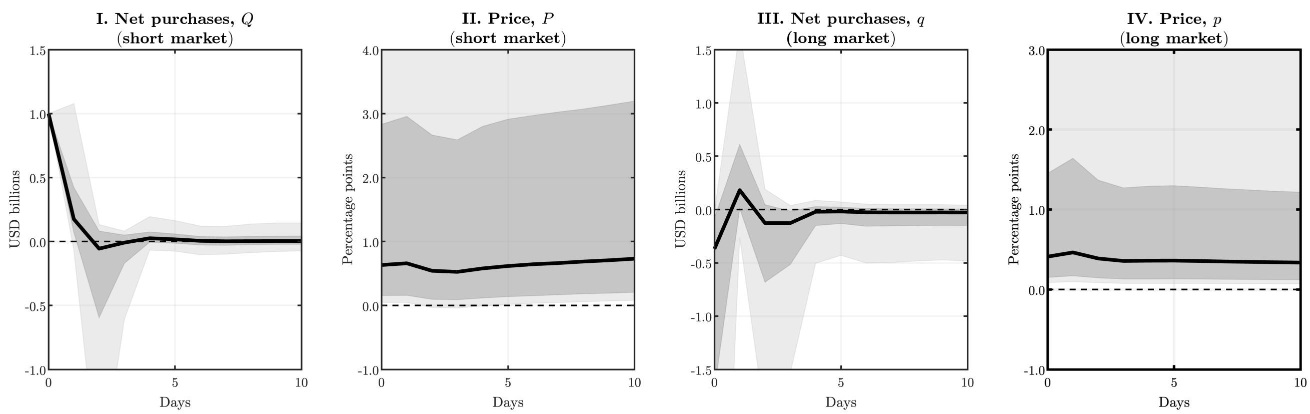
Note: Panel (a) presents impulse response functions to a dealer supply shock scaled to lower quantities in the short market (i.e. purchases of inflation protection by hedge funds) by \$1bn. Panel (b) presents impulse response functions to a pension fund demand shock scaled to raise quantities in the long market (i.e. purchases of inflation protection by pension funds) by \$1bn. Panel (c) presents impulse response functions to a hedge fund demand shock scaled to raise quantities in the short market by \$1bn. Across all panels, these demand and supply shocks are identified using the sign restrictions identification strategy. The posterior is simulated using 10 million draws using the sampler of Arias, Rubio-Ramírez, and Waggoner (2018) that selects 14,922 importance sampling draws for inference. The bold line indicates the median of the draws from the sampler and shaded bands are its 68% and 90% credible intervals.

Figure 15 IMPULSE RESPONSE FUNCTIONS TO AN INFLATION SHOCK: ESTIMATES FROM THE SIGN RESTRICTIONS IDENTIFICATION STRATEGY

(a) Baseline result

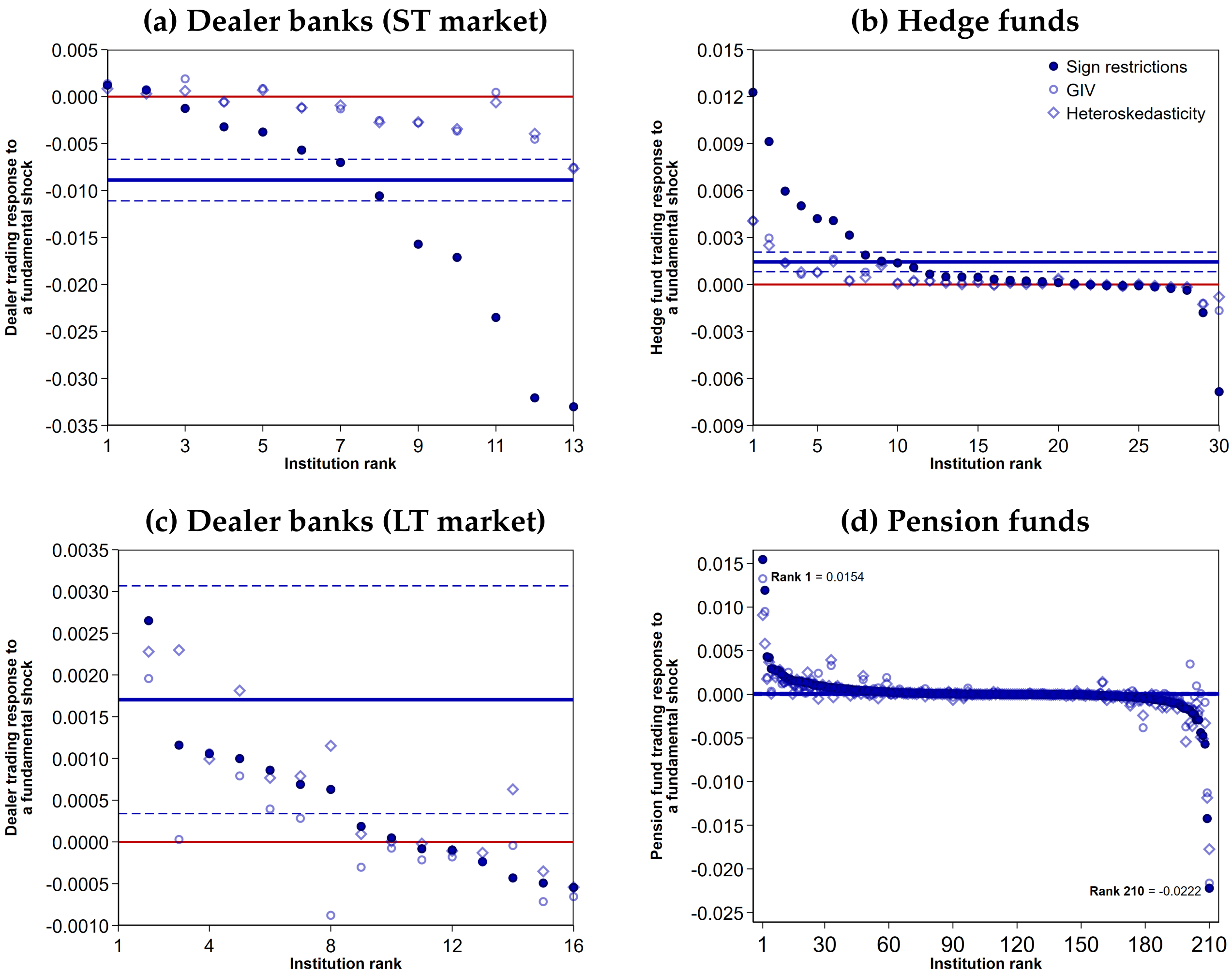


(b) Adding insurers



NOTE: Impulse response functions to a inflation shock scaled to raise quantities in the short market (i.e. purchases of inflation protection by hedge funds) by \$1bn. Panel (a) presents impulse response functions to an inflation shock estimated from the sign restrictions identification strategy without including insurers in the estimation data (which is the baseline result presented in the main text). Panel (a) presents impulse response functions to an inflation shock estimated from the sign restrictions identification strategy that includes insurers in the estimation data. Across all panels, we present the median of the draws from the sampler along with 68% and 90% Bayesian credible intervals.

Figure 16 ESTIMATES OF β_i FOR INDIVIDUAL INSTITUTIONS



NOTE: Individual markers denote point estimates of β_i across the three different strategies. Institutions ranked based on the sign restrictions strategy. The thick blue lines are pooled coefficient estimates estimated by three-stage least squares, using our identified inflation shocks as an instrument for the change in the long-maturity inflation swap breakeven rates. The dashed blue lines are their 95% credible sets, with standard errors clustered at the institutional-level (where each is identified by a legal entity identifier). For figures (d), the respective responses of the highest and lowest ranking institutions are not plotted and their values are instead indicated on the figures as shown. Sample goes from January 2019 to February 2023.

References

Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2018): "Inference Based on Structural Vector Autoregressions Identified With Sign and Zero Restrictions: Theory and Applications," *Econometrica*, 86, 685–720.

Bahaj, S. (2020): "Sovereign Spreads in the Euro Area: Cross Border Transmission and Macroeconomic Implications," *Journal of Monetary Economics*, 110, 116–135.