The Market for Inflation Risk☆

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

January 2024

Abstract

This paper uses transaction-level data on UK inflation swaps to characterize who buys and sells inflation risk, when, and with what price elasticity. This provides measures of expected inflation cleaned from liquidity frictions, and of the varying influences of market participants with different beliefs. We first show that this market is segmented: pension funds trade at long horizons while hedge funds trade at short horizons, with dealer banks as their counterparties in both markets. This segmentation suggests three identification strategies—sign restrictions, granular instrumental variables, and heteroskedasticity—for the demand and supply functions of each investor type. We find that swap prices absorb new information quickly, the supply of long-horizon inflation protection is very elastic, short-horizon price movements are unreliable measures of expected inflation as they primarily reflect liquidity shocks, and that long-horizon price movements overstates changes in expected inflation during important events if they were not cleaned from liquidity shocks.

Keywords: asset demand system, monetary policy, anchored expectations, identification of demand and supply shocks.

JEL Codes: E31, E44, G12, C30.

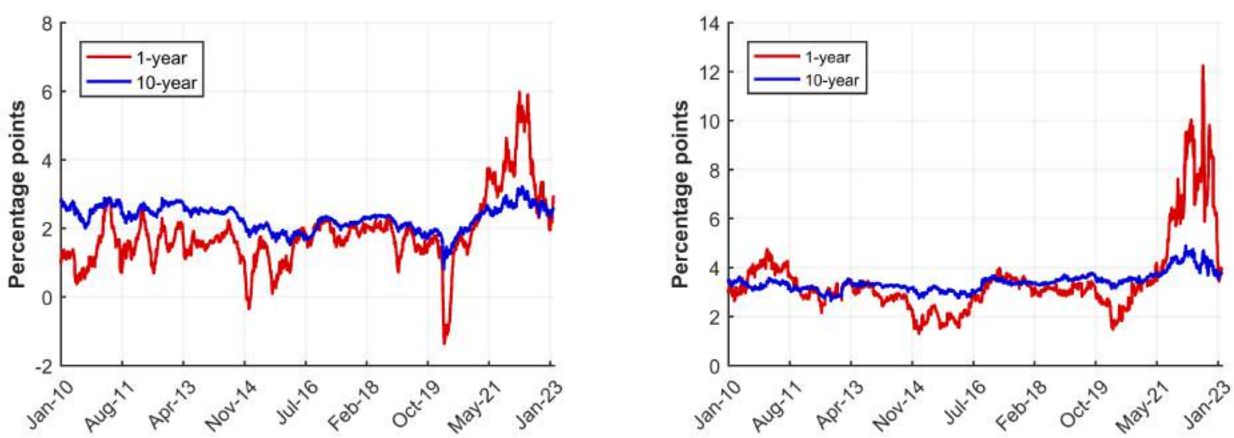
☆The authors would like to thank seminar and conference audiences at the Bank of England, City University Hong Kong, Princeton University, University of Cambridge, University of Kent, University of Liverpool, University of York, Hong Kong Baptist University, the NBER Summer Institute, the 3rd Qatar Centre for Global Banking and Finance Annual Conference, the 15th Annual Paul Woolley Centre Conference, the CESifo Venice Summer Institute, as well as Carolin Pflueger, Gyuri Venter, Aytek Malkhozov, Nina Boyarchenko, Markus Brunnermeier, Andrew Hauser, Misa Tanaka, Dimitri Vayanos, and Motohiro Yogo for their comments. E-mail contacts: `s.bahaj@ucl.ac.uk`, `robert.czech@bankofengland.co.uk`, `s.ding5@lse.ac.uk`, and `r.a.reis@lse.ac.uk`. The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England or its committees. Reis is grateful to UKRI grant 109166 for financial support.

1 Introduction

A large and fast-growing market for inflation swaps allows investors to trade inflation risk. Prices in this market are closely watched because they provide a measure of what market participants expect inflation will be that complements (and in some ways is superior to) the measure from inflation-indexed government bonds (the break-even rates). Scores of papers in monetary policy have used these prices to report high-frequency measures of expected inflation at multiple horizons and to study how they respond to identified shocks (see, e.g., Haubrich, Pennacchi, and Ritchken, 2012; Beechey, Johannsen, and Levin, 2011; Andrade and Ferroni, 2021). Speeches by policymakers often show plots of these prices over time, especially of forward swaps at long horizons, to assess whether monetary policy is credibly delivering price stability (see, e.g., Mann, 2022; Ramsden, 2022). More recently, the large gap between 1-year and 10-year measures of expected inflation in the US and the UK shown in Figure 1 has been interpreted as revealing that markets believe that current high inflation will be temporary and that expectations have stayed anchored.

Figure 1 Prices of inflation swaps in the recent past

(a) United States **(b) United Kingdom**



Note: Prices shown are zero-coupon swap breakeven rates. Source: Bloomberg.

This paper studies the relevant players and quantities behind these prices. Using detailed regulatory transaction-level data on every over-the-counter (OTC) inflation swap contract sold in the UK, we identify who buys and sells inflation insurance on a daily basis and study how the prices are formed. With a model of demand and supply, we decompose the price movements into fundamentals and liquidity shocks. From the joint variation in prices and quantities, we learn which shocks drive prices at different dates, and about the slopes of demand and supply curves for inflation protection by different investor types. With this who and what, we then revisit the account of how inflation expectations moved during the pandemic, the subsequent period of increasing inflation, and the UK LDI crisis of September and October of 2022. Finally, we decompose whose beliefs drove expected inflation across and within sectors.

Contributions and outline: We provide three contributions. The first is a description of a large and important financial market that is somewhat opaque because of its over-the-counter nature. We find that this market fits remarkably well into a segmented-markets model. Namely, Section 2 establishes three stylized facts on the UK market for inflation swaps. First, dealers are not neutral market makers, but they are net sellers of inflation protection, beyond their holdings of index-linked gilts. Moreover, increases in the sale of swaps are only weakly correlated with changes in the holdings of indexed bonds. Second, at long horizons (10 years or more) the buyers of the swaps are mostly pension funds. They hold large, persistent, positive net positions, and much of the variation in sales and trading volume is driven by the actions of pension funds. Thus, pension funds are net receivers of inflation in this market while dealers are the net payers: when inflation rises unexpectedly, there is a direct flow of payments from dealers to pension funds. Third, at short horizons (3 years or less), informed traders such as hedge funds hold small net positions, but actively trade, so that in any given day they can have a negative or positive net position. When short-term inflation rises, sometimes they win, sometimes they lose, and on the other side, dealer banks likewise have fluctuating net positions. In other words, we observe a remarkable *segmentation* of this market: pension funds barely trade in the short-horizon market while informed traders conduct most of their buying and selling activity in the short-horizon market.

Motivated by these facts, Section 3 provides a model of the market for inflation risk with two main characteristics. First, it is a portfolio choice model, as opposed to a model of broker-dealers, since dealers in this market hold persistent large net positions. Second, it features two markets, at short or long horizons, where inflation risk is demanded by two separate agents — hedge funds and pension funds — with one type of agents — dealer banks — supplying the inflation swap contracts. The model provides a decomposition of inflation swap prices into expected inflation, compensation for risk, and liquidity frictions, and identifies the primitives and market frictions that can give rise to them.

The model allow us to state clearly the assumptions behind our second contribution: empirical identification strategies for segmented market models. Section 4 describes three identification strategies to estimate the demand system for inflation protection across institutions. The first strategy exploits the high frequency of the data. It assumes that, within a day, hedge funds respond more to fundamentals than dealer banks, and dealer banks respond more than pension funds. These assumptions impose restrictions on the relative shifts of demand and supply functions in response to shocks that amount to sign restrictions on the structural responses of prices and quantities to various shocks. The second strategy exploits the cross-sectional variation in the daily trading activities of each institution. This transaction-level data is highly granular, with some institutions taking larger positions that fit a power law distribution. We construct a granular instrumental variable, using institution-level disturbances as instruments for the aggregate demand for each class of agents. The final identification strategy exploits the heteroskedasticity in our time series data that spans nearly four years. Releases of inflation data are regular and cause heightened volatility in the inflation swap market. Assuming that shocks to fundamentals drive most of the heteroskedasticity on these dates, this time-varying volatility identifies these shocks.

Sections 5-7 present our third and final contribution, a set of estimates that are consistent across our three identification strategies. These estimates are split into three separate applications on: how this financial market works, how to accurately measure expected inflation, and whose beliefs drive those expectations. Starting with the first, we find that: (i) Impulse responses to shocks become horizontal within one to three days. The inflation swap market seems to incorporate new information relatively quickly. (ii) The supply of inflation protection by dealer banks to pension funds at long horizons is very elastic, unlike their supply to hedge funds at short horizons. (iii) Most of the variation in the price of short-horizon inflation swaps is driven by liquidity shocks, and only about one third of those arise from the supply by dealer banks, with the majority being instead shocks to demand.

Second, we provide a novel time series for long horizon expected inflation that is cleaned of liquidity frictions between January 2019 and February 2023. Our estimates suggest that an unfiltered reading of inflation swap prices will lead to an overstatement of movements in expected inflation. During the pandemic, the uncleaned conventional measures overstated the risk of deflation, and during the energy crisis they overstated the risk of inflation, while from September 2022 to February 2023 they have overstated the extent to which expectations were unanchored.

Finally, we decompose the movements in expected inflation into their individual components across agents. We find that there is wide disagreement among dealer banks, and among hedge funds about expected inflation. However in the short horizon market, three dealer banks and three hedge funds, on account of their size and extreme views, have a large price impact relative to all others. In the long horizon market instead, the price impact of individual pension funds is similar and smaller. Strikingly we find that our model's implied ranking of optimistic and pessimists from market activity among dealer banks matches well the answers to surveys by the same institutions.

Connection to the literature: Our paper is related to five strands of the literature. First, we contribute to the literature on segmented markets that builds on Vayanos and Vila (2021). We highlight a particular market where the segmentation across horizons is clear. While this market for inflation risk is not as large as the markets for bonds and foreign exchange that the prior literature has focused on, it it is still significant, and it has the virtue that we have transaction-level data on a large sample of the market. Therefore, we can cleanly empirically identify the arbitrageurs and preferred-habitat agents that are described in those theories. We provide three complementary identification strategies that may be useful in other estimations of segmented market models.

Second, we estimate an asset demand system, in the footsteps of Koijen and Yogo (2019) or more recently Gabaix, Koijen, Mainardi, Oh, and Yogo (2022). While previous studies have used data on stocks (e.g., Koijen, Richmond, and Yogo, 2020), bonds (e.g., Koijen, Koulischer, Nguyen, and Yogo, 2021) or exchange rates (e.g., Koijen and Yogo, 2020), we focus on the large and liquid over-the-counter market for inflation swaps. Our approaches to identification are different, as is our focus on extracting measures of expected inflation from the prices of swaps. Within this literature, in our use of granular instrumental variables (Gabaix and Koijen, 2020), we are joined by Gabaix and Koijen (2021). Differently, we use trade-level positions across different investor types to build a granular instrumental variable for the demand of each type.

Third, and related to the previous two, Begenau, Piazzesi, and Schneider (2015), Hanson, Malkhozov, and Venter (2022), Jiang, Matvos, Piskorski, and Seru (2023) and McPhail, Schnabl, and Tuckman (2023) focus on the interest rate swap market to uncover the types of investors who bear interest rate risk and how their exposure to risk varies. Begenau, Piazzesi, and Schneider (2015), Jiang, Matvos, Piskorski, and Seru (2023) and McPhail, Schnabl, and Tuckman (2023) estimate the positions taken by banks, while Hanson, Malkhozov, and Venter (2022) assume a risk profile for the intermediating dealers as a proxy for their positions. Instead, we focus on inflation derivatives, and we directly observe the directional positions taken on by dealer banks and other investors. Our results, in contrast to the evidence by Jiang, Matvos, Piskorski, and Seru (2023) and McPhail, Schnabl, and Tuckman (2023) on interest rate swaps, show that banks do take large net positions in the inflation risk market. We also find that liquidity shocks affecting the clients of dealers play an important role in driving market prices alongside shocks to dealers themselves. Like Hanson, Malkhozov, and Venter (2022), we use an affine representation of the structural shocks and, in one of our identification strategies, we use sign restrictions, but we further complement this with other strategies so we can cross-validate the results, all pointing to the same conclusions on what drives the prices of inflation swaps.

Fourth, our focus on inflation risk and how it is priced by financial markets is shared with a long literature: see Cieslak and Pflueger (2023) and D'Amico and King (2023) for recent surveys.[1] A perennial concern in this literature is whether market prices reflect not just subjective expectations and compensation for risk, but also liquidity premia, a catch-all term for market imperfections that can be large and vary over time. We address this challenge head on, contributing to a better understanding of where these premia come from, and providing superior estimates of market-expected inflation. Also, as in Reis (2020), we contrast market behavior with survey measures of expected inflation, but we have micro data and estimate to map this to institution's behavior as

[1] Recent contributions include, among others, Campbell, Pflueger, and Viceira (2020), Boons, Duarte, de Roon, and Szymanowska (2020) and Fang, Liu, and Roussanov (2022).

opposed to just price outcomes.

Fifth, and finally, we use the regulatory EMIR Trade Repository (TR) data on trade-level derivative positions to shed light on OTC derivative markets (Abad et al., 2016). Cenedese, Della Corte, and Wang (2021) and Hau, Hoffmann, Langfield, and Timmer (2021) use the EMIR TR data on FX forwards and swaps to investigate the impact of leverage ratios and price discrimination, while Cenedese, Ranaldo, and Vasios (2020) use data on the interest rate swap market to compare prices in over-the-counter transactions versus prices in centrally-cleared trades.[2] To the best of our knowledge, our work is the first to use the granular EMIR TR data on inflation swaps, which enables us to reveal who is buying the swaps, who is selling them, and the associated gross and net positions.

Outline: This paper is organized as follows. Section 2 describes the granular transaction-level data and the set of novel stylized facts that characterize the UK inflation swap market. Section 3 presents the model and a formal characterization of the asset demand system. Section 4 discusses the set of three identification strategies and their empirical implementation. Sections 5-7 presents our estimates split into three sections. The first discusses the estimates on how this financial market functions: how quickly it absorbs information, what shocks drive it, and what are the slopes of demand and supply. The second discusses the implications for macroeconomic measures of expected inflation, and how swap prices can overestimate these movements. The third discusses the estimates on whose beliefs matter the most in these measures of expected inflation and compares the results with those from surveys. Section 8 concludes.

2 Data, summary statistics, and stylized facts

We start by describing the market, the source of our novel data, and some summary statistics for the UK inflation swap market. Then, we establish three main facts.

2.1 Inflation swap contracts

The inflation swap market is an over-the-counter (OTC) market, where the terms of a transaction are negotiated privately between the two counterparties involved in the trade. Most contracts take the form of zero coupon swaps linked to an index measure of inflation. This is a contract where cash changes hands at the end of the swap contract on a legally binding settlement date. The floating rate payer pays one plus the total growth of the inflation index over the life of the contract times the gross notional. The fixed rate payer pays the compounded fixed rate times the

[2] Other recent contributions include, for instance, Benos, Ferrara, and Ranaldo (2022), Czech, Della Corte, Huang, and Wang (2022) and Ferrara, Mueller, Viswanath-Natraj, and Wang (2022).

notional. The fixed rate of the swap contract is set ex ante such that the net present value payoff of the swap to both counterparties is equal to zero at initiation: for this reason, it is commonly referred to as the swap breakeven rate. The realization of the floating inflation is uncertain, and so the liability associated with this leg is determined ex post.

There is inflation risk since, when realized inflation rate deviates from the fixed rate, one counterparty becomes liable to pay insurance to the other counterparty through a net cash flow payment based on the contract's notional amount. The inflation measure for the UK inflation swap market is the retail price index (RPI), as this dominates nearly all swap contracts traded on UK inflation. This is partly because the RPI index is the index used for inflation-linked gilts. In the past decade, the RPI tends to be approximately 1.5 percentage points or more above the CPI, which is the base of the 2% target of the Bank of England.[3]

2.2 The EMIR Trade Repository Data

During the G20 summit in September 2009, it was agreed that derivative trades should be reported to trade repositories, thus granting regulators access to high-quality and high-frequency data. In the UK, the effort to increase the transparency of derivative markets was part of the European Market Infrastructure Regulation (EMIR), which makes it mandatory for UK legal entities to report the terms of any derivative transaction to a trade repository authorized by the Financial Conduct Authority (FCA) by the next business day.[4]

We rely on the EMIR Trade Repository (TR) data to obtain trade-level information on OTC inflation swaps. Our data sample spans the period from January 2019 to February 2023, and it consists of all trades submitted to DTCC Derivatives Repository Plc ('DTCC') in which at least one of the counterparties is a UK-regulated entity. Since this is the largest trade repository in terms of market share, we are confident that it captures a volume of derivatives trading data that is highly representative of the market.[5] We use the DTCC's daily trade state reports to capture all outstanding inflation swap trades on a given day, as well as to obtain the flow of trading activities. We allocate investors to groups using a best-endeavor sectoral classification.

[3] Appendix E.1 further discusses the comparison between price indices for the UK and Appendix E.6 describes the other indices used in inflation swap contracts in our data. 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.

[4] Since the UK's departure from the European Union, the EMIR has been adapted into UK legislation under UK-EMIR. As of February 2023, there are four trade repositories regulated by the Financial Conduct Authority that are registered under UK-EMIR to operate in the UK. These are: DTCC Derivatives Repository Plc, UnaVista Limited, REGIS-TR UK Limited and ICE Trade Vault Europe Limited.

[5] In Section B, using an alternative data source for insurers' swap holdings, we find that the DTCC trade repository captures the vast majority of client trades.

The raw data has approximately 3.5 billion observations. As expected, it is noisy. We identify the inflation swap contracts from all other interest rate derivatives and clean the data by eliminating: (i) duplicated transaction-level reports at the counterparty-trade ID-other counterparty level, (ii) intragroup transactions, (iii) compression trades, (iv) trades with implausible notional amounts (greater than \$10bn and lower than \$1000), and (v) trade reports that do not meet the set of UK EMIR validation rules. Finally, we drop all observations with inconsistent values for either the reported notional, the identities of the counterparties, the counterparty side, the maturity date, or the underlying inflation index. Finally, since the vast majority of trades are spot contracts (rather than forward contracts), we focus on this predominant segment of the inflation swap market. This leaves us with more than 25 million observations, on a daily basis, from 2^{nd} January 2019 to 10^{th} February 2023.[6] We provide further details on trade repository data and our data cleaning procedure in Section A.1 of the Appendix.

Appendix B compares our trade repository dataset with supervisory data over some institutions that has detailed information on the derivatives holdings every quarter. We find that the holdings match quite closely our data.

2.3 The size of the market and the dealer-client segment

The left panel of Figure 2 shows the average gross notional amount outstanding across the sample. The market size is substantial, varying between \$3.5tn and \$4tn, which is roughly 110-130% of UK GDP. By comparison, outstanding index-linked government bonds were approximately 20% of GDP during the same period.

Approximately 60% of the total gross notional amount outstanding in the market is in the centrally-cleared market segment (CCP). Only clearing members, which are mostly dealer banks, have access to this market. Therefore, data from clearing houses do not allow for the study of which types of client institutions demand and trade these contracts. We instead have data on the dealer-client segment of the market that is not centrally cleared. Still, 17% of the trades are between dealers, and so suffer from the same problem, so we exclude these.

Instead, we focus on remaining 22% of contracts that are between dealers and clients.[7] These contracts are sold by 16 large international banks (dealers for short) and bought by different financial institutions, including pension funds and liability driven investors (PFLDIs, or just pension funds, for short) as well as informed traders (like hedge funds) and others (including non-dealer

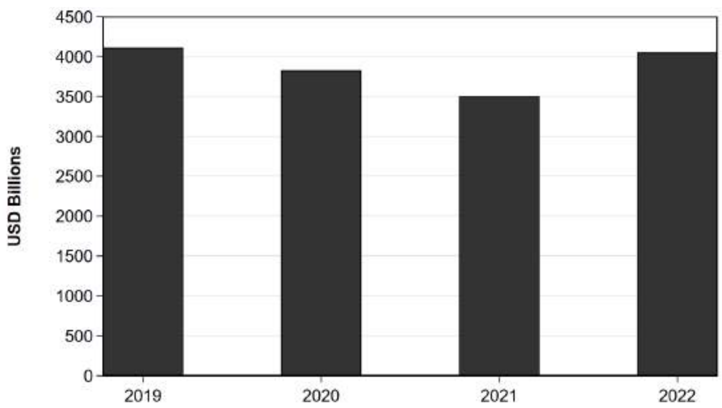
[6]Section A.3 of the Appendix describes the variables more precisely.

[7]In our data, transactions between a clearing member and the CCP appear as a unique trade in the dataset, although they may represent one half of a trade with another clearing member, as well as one or multiple trades between the clearing member and her client. Another benefit of focusing on the dealer-client market is that we avoid double-counting.

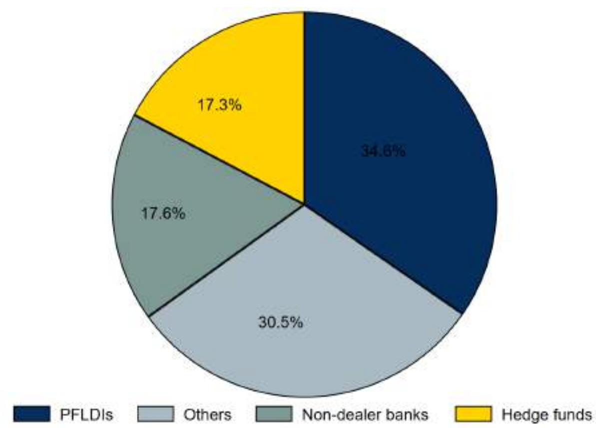
banks, insurance companies, and no-financials, among others).[8] The right panel of Figure 2 shows the main sectors of the client side of the market.

Figure 2 Gross notional amounts: average and compositions

(a) Average volume outstanding



(b) Average market shares



Note: The left figure shows the stock of inflation swap contracts outstanding measured by gross notional amount traded by all investors in the market in a given year, averaged across monthly trade state reports at month-end. The right figure shows the distribution of total gross notional traded by various client institutions against dealers in the dealer-client segment of the market, computed as an average across all monthly trade state files Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

2.4 The clients by maturities

Figure 3 plots the gross notional positions outstanding in the dealer-client segment of the inflation swap market since 2019, split by maturity and by client sector. These have grown rapidly, reaching a peak of around \$1.1tn in late 2022, as more agents have bought and sold protection against inflation.

Across horizons, the larger share of the market is in contracts for inflation for a long horizon of ten years or more, but there is also a significant market at a short horizon, from the present to three years away.

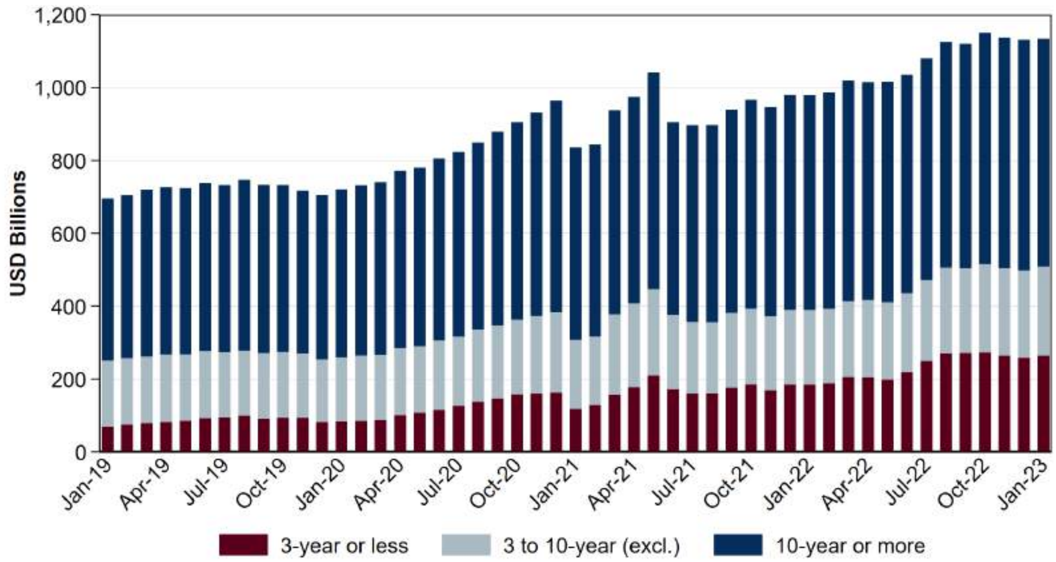
Across clients, hedge funds have steadily increased their notionals since the COVID-19 market turmoil in 2020 and the reappearance of inflation in 2021, from less than \$50bn in 2019 to around \$200bn in 2022.[9]

[8] The full list of investor types also includes asset managers, sovereign wealth funds, trading services, proprietary trading firms, central banks, state and supranational institutions, but the trading volumes of these groups tend to be small.

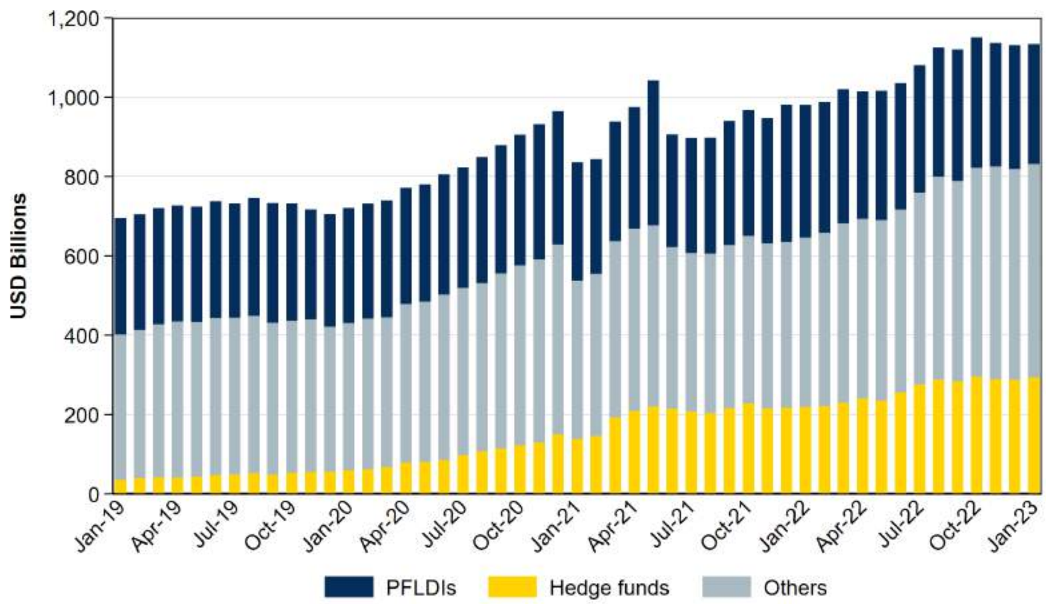
[9] Appendix A.4 shows a few more summary statistics describing the composition of the data.

Figure 3 THE QUANTITIES AND THE INSTITUTIONS BEHIND THE PRICES

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



(b) Gross notional by client sector



NOTE: Gross notional positions outstanding that are traded in the dealer-client segment of the market across all price indices. SOURCE: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

2.5 Three facts on the segmented UK inflation swap market

To establish the main facts, we focus on *net* notional amounts. These are defined as the sum of net protection bought (sold) by counterparties that are net buyers (sellers) of inflation swaps.

The first fact is that *dealer banks are not neutral market makers*. Rather, as illustrated by the top panel of Figure 4, dealer banks have issued an amount of inflation protection in this market that is beyond their holdings of index-linked government bonds (the natural hedge against their inflation risk exposures) since at least the second half of 2019.[10] The magnitude of their exposures is substantial and reached a peak of almost \$150bn in 2020.[11] Rather than being driven by the risk-taking behavior of a few large dealer banks, this phenomenon is consistent across the entire sector: nearly all dealer banks sell inflation protection and thereby take on inflation risk.[12] Moreover, Figure E.2 in the Appendix shows that the quarterly change in the holdings of index-linked government bonds of individual dealer banks is only weakly correlated with the change in their net notionals in UK RPI inflation swaps (with a correlation coefficient of -0.23).

The bottom panel of Figure 4 reveals that it is primarily PFLDIs that take the opposite position to dealers. They have persistently large and positive net notional positions in this market, with a maximum of approximately \$100bn in our sample. Given their gross notionals of around \$250bn in 2022 (see Figure 3), the PFLDI's high net-to-gross ratio emphasizes the largely one-directional appetite of PFLDIs for buying inflation protection.[13] Hedge funds, in contrast to PFLDIs, tend to switch between being net buyers and sellers of inflation swaps, consistent with prior studies documenting the role of hedge funds as arbitrageurs and informed traders in derivative markets (see, e.g., Czech, Della Corte, Huang, and Wang, 2022).[14]

The second fact is that *PFLDIs are buyers of inflation protection mainly at long horizons.* Figure 5 shows the net notional positions of both PFLDIs and dealers broken down by the initial time-to-maturity of the contracts. PFLDIs hold a sizable amount of inflation swap contracts of 10-year initial maturity or longer, relative to their overall net position. This persists over time, and is consistent with PFLDIs seeking to buy inflation protection to hedge their long-dated liabilities.[15] Dealers make a market to meet this demand, and persistently hold a sizable amount of inflation

[10] To obtain data on the index-linked gilt holdings of dealer banks, we use the UK banking system's Global Network of granular exposures, which captures roughly 90% of the UK banking system's total assets. The data and cleaning procedure is described in Covi, Brookes, and Raja (2022).

[11] Figure E.5 shows that the average maturity of index-linked government bonds held by dealer banks is relatively well aligned with the average maturity of their outstanding UK RPI swap positions. The average swap maturity is remarkably stable at around ten years.

[12] Bearing in mind the limitations of our data coverage in the EU jurisdiction as a consequence of Britain's decision to leave the EU, we also calculate in Appendix E.9 the net notional positions of the same client types in the dealer-client segment in the EU inflation swap market. It is also true in this market that pension funds are net buyers of inflation protection, hedge funds switching sides from being a net buyer to a net seller from time to time, and dealer banks trading as their counterparties that are net sellers.

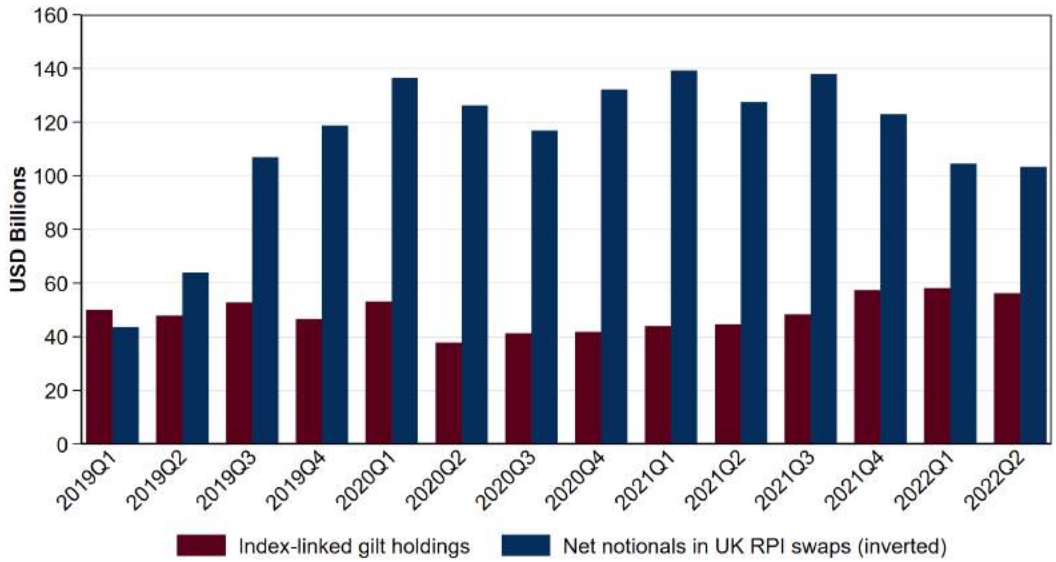
[13] The pronounced appetite of pension funds for inflation protection may be driven by the limited supply of index-linked gilts, which cover only a small fraction of pension funds' total liabilities, as shown in section E.3 of the Appendix.

[14] For comparison, as of February 2022, in the UK interest rate swap market total net exposures amounted to \$500bn (Khetan et al., 2023). The corresponding number for the UK RPI inflation swap market in February 2022 is around \$125bn, hence approximately one-fourth the size of the interest rate swap market in terms of net exposures.

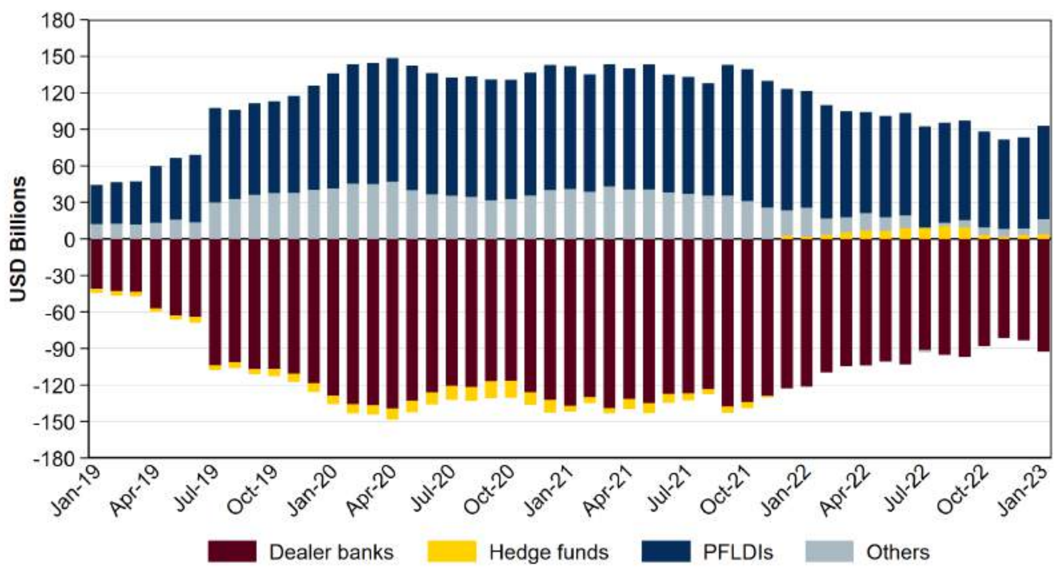
[15] Figure E.4 shows that the approximated average maturity of index-linked government bonds held by pension funds is around six to seven years longer the maturity of their outstanding UK RPI swap contracts, likely due to the extremely long maturities (often more than thirty years) of recent issuances of index-linked gilts by the UK Debt Management Office (DMO).

Figure 4 Dealers have a non-zero net exposure to inflation risk

(a) Dealer banks' net notional position v.s. index-linked gilt holdings



(b) Net notional position by investor type in the UK RPI inflation swap market



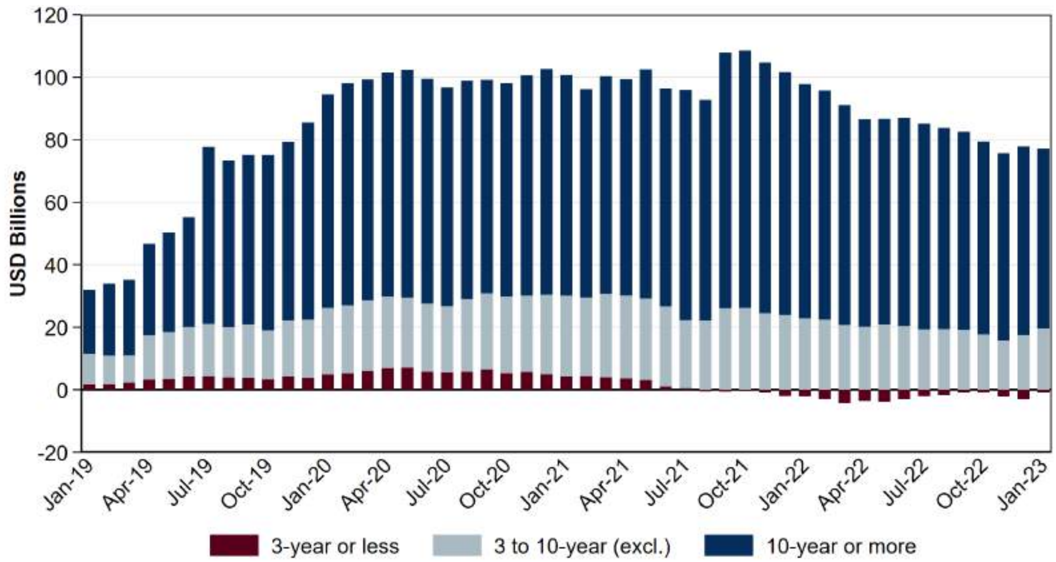
Source: UK banking system's Global Network of granular exposures and DTCC Trade Repository OTC interest rate trade state files, from March 2019 to June 2022.

risk in this segment.

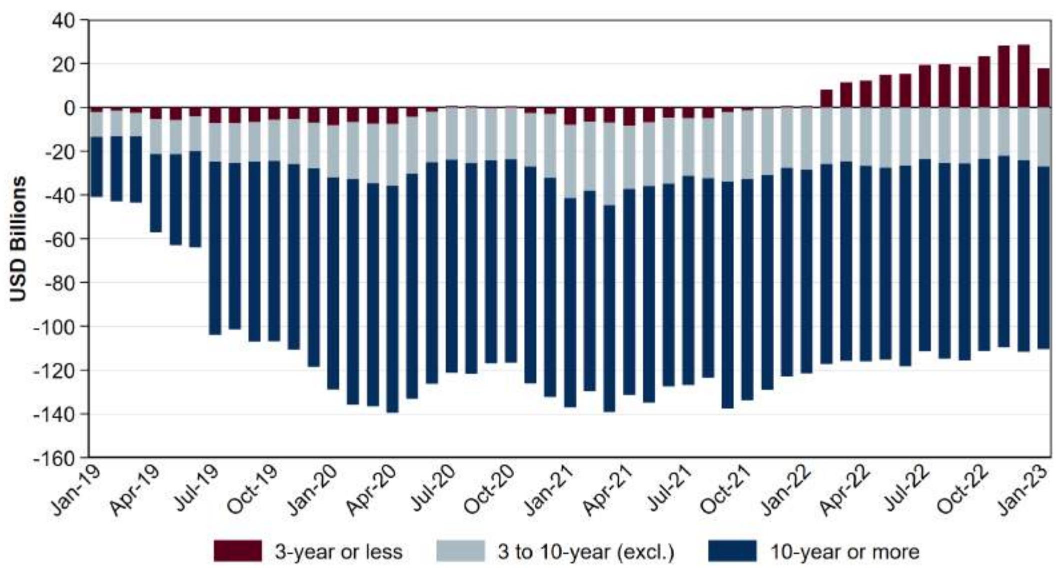
The third fact is that *hedge funds trade inflation risk primarily in the short horizon market.* Figure 6 shows the net notional position taken on by hedge funds. They mostly trade inflation swaps with an initial maturity of 3 years or less. Further, they switch between being net buyers and net sellers, which is consistent a priori with them following informed trading strategies that try to exploit arbitrage opportunities.

Figure 5 Pension funds buy protection from dealers in the long horizon market

(a) Net notional position of pension funds in the UK RPI inflation swap market



(b) Net notional position of dealer banks in the UK RPI inflation swap market



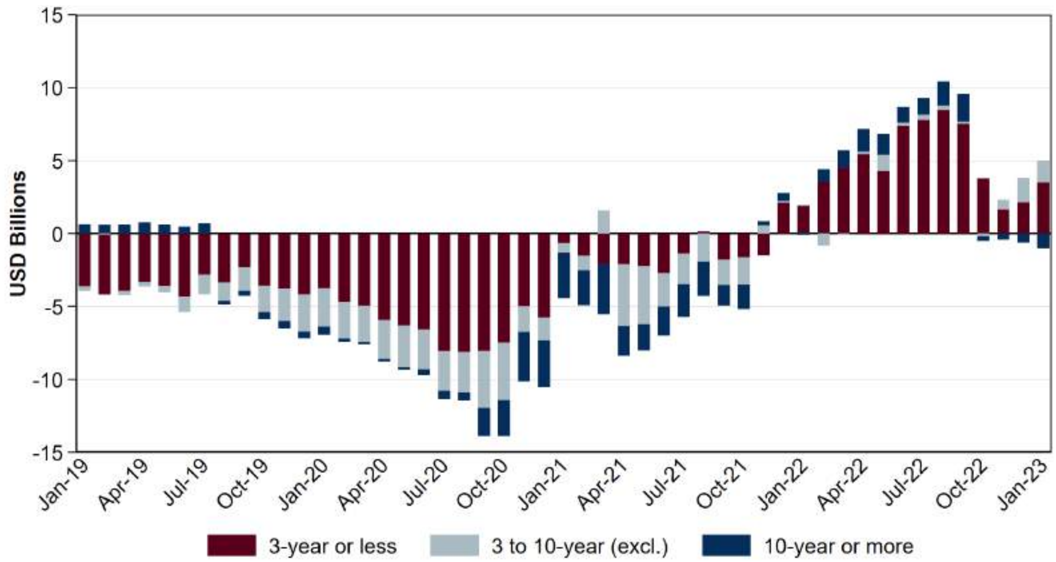
Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

Combined, these three facts imply a remarkable *segmentation* of the UK RPI market: PFLDI institutions primarily trade in the long horizon market where they hold persistently large positive net positions, hedge funds trade in the short horizon market with fluctuating net positions, and dealer banks are the counterparties in both markets to both types of clients, so they trade actively in both. Dealers fit into the category of arbitrageurs across maturities, while PFLDIs and hedge funds fit into the preferred-habitat investor category in the spirit of Vayanos and Vila (2021).

Looking for more evidence of this market segmentation, we utilize the high frequency of the

Figure 6 Hedge funds are active in the short horizon market

Net notional position of hedge funds in the UK RPI inflation swap market



Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

data. The top panel of Figure 7 shows the gross notional positions aggregated at the sectoral level at a daily frequency. Clearly, PFLDI's gross trading volume in the long horizon market dominates that from the hedge funds, and vice versa in the short horizon market. There is some trading activity of hedge funds in the long horizon market, and of PFLDIs in the short horizon market, but these are small relative to the predominant type in the market, or relative to their large positions in the other market.[16]

The bottom panel of Figure 7 instead calculates the median maturity of trades from both the PFLDIs and hedge funds. The segmentation of the markets is also clear here.

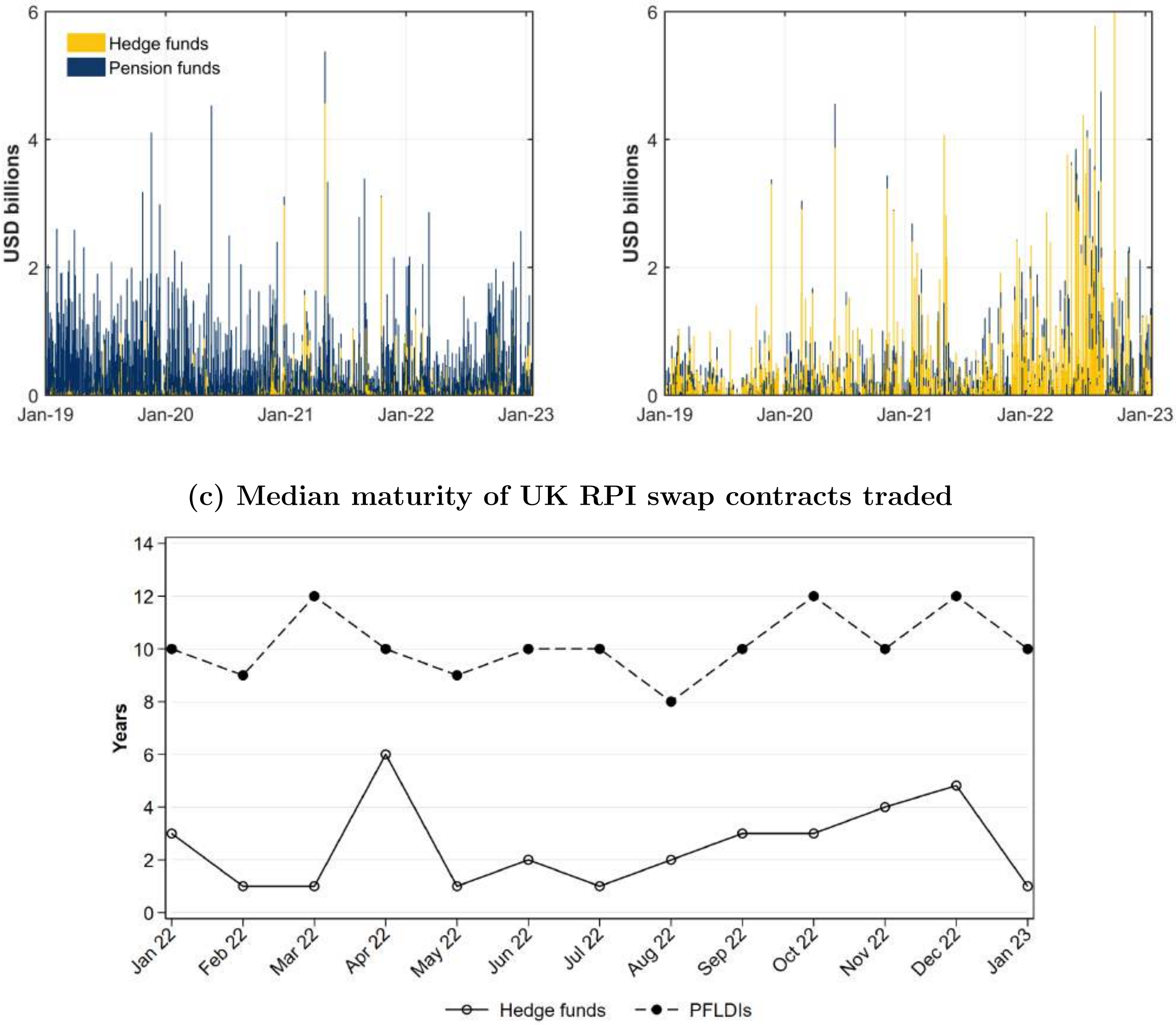
[16]Appendix E.7 further plots gross notional positions and amounts traded across horizons by investor type. Even by this measure, PFLDIs are mostly active in the long horizon market, while hedge funds trade large amounts in the short horizon market. Hedge funds sometimes take positions in the long horizon market, but these are almost always small compared to their positions in the short horizon market.

Figure 7 Market segmentation at higher frequencies

(a) Daily volume: long horizon market

(b) Daily volume: short horizon market

(c) Median maturity of UK RPI swap contracts traded



Note: The top panel plots the gross notional traded by both the hedge fund and PFLDI sectors at a daily frequency for a given execution date of the trade. This is aggregated from contract-level data pertaining to each institution belonging to both sectors. The bottom panel shows the volume-weighted median maturity of the executed trades by both sectors for the most recent period in our sample. Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

3 A model of the demand for inflation risk

A model that fits our facts should include three types of agents where all, including dealers, take non-zero net positions and hold on to them, and (at least) two separate markets, for both short-horizon and long-horizon inflation risk, with hedge funds in the first and pension funds in the latter, while dealers are active in both. This section offers such a model. The model imposes two main assumptions, justified by the stylized facts, as well as minor assumptions on functional forms and uncertainty whose only purpose is to lead to analytical linear demand functions. The goal is not to provide a structural representation to directly estimate in the data, but rather to have a vehicle to clarify concepts, catalog different types of shocks, and later state clearly the identification assumptions.

3.1 The optimization problems

There are three types of agents: two clientele traders that are the pension funds (f) and hedge funds (h), and a dealer bank sector (b) acting as their counterparties. There are many institutions within each group, indexed by i. The objective of each institution (and taking pension funds for concreteness) is to solve a static portfolio allocation problem where they maximize utility of terminal wealth $a'_{f,i}$ according to a constant absolute risk-aversion utility function:

$$\mathbb{E}_{f,i}\Big[-\exp\left(-\widetilde{\gamma}_{f,i}a'_{f,i}\right)\Big] \tag{1}$$

The Arrow-Pratt coefficient $\widetilde{\gamma}_{f,i} = \gamma_{f,i}/a_{f,i}$, so that it can scale with initial wealth $a_{f,i}$; otherwise, larger institutions would mechanically hold riskier positions.

We interpret the single period in our model as a trading day. An inflation swap contract for the long horizon is an asset that costs a fixed coupon p at the end of the day, and pays off π. This is not the realization of inflation per se, but rather the realization of the value of the floating portion of the swap contract at the end of the day. Swap contracts are subject to variation margin so that counterparties exchange cash to ensure that the contract's net value remains at zero at the end of the period, so these payments can be interpreted as meeting a margin requirement. An equally valid interpretation is that the position is closed down at day-end with the agent whose side of contract appreciated receiving a payment.

Normalizing the daily real safe rate to 1, the budget constraint of a pension fund is:

$$a'_{f,i} = a_{f,i} + (\pi - p)q_{f,i} + (d - s)e_{f,i} + y_{f,i} \tag{2}$$

Aside from swaps, the agent also trades a market asset that costs s and pays d. This could be the result of a portfolio choice involving many other assets, but as we will focus solely on the demand for inflation risk, taking $e_{f,i}$ to be the total amount invested in those assets, we do not need to solve for its components. Finally, $y_{f,i}$ is a background risk that cannot be traded due to incomplete markets, and which may be correlated with inflation, such as the payments to pension holders or the inflows into new pensions.[17]

[17] The presence of these two sources of incomes that maybe correlated with inflation also clarifies that the net notionals $q_{f,i}$ that we measure in the previous section do not correspond to the total exposure of an institution to inflation risk. That total exposure depends also on hos these other sources of income vary with inflation.

Let $\Theta_f \in \mathbb{N}$ denote the set of institutions in the pension fund sector. Within this set, each institution has individual beliefs captured in the expectations operator $\mathbb{E}_{f,i}(.)$. For the expectations of asset payoffs and background risk, we assume they all agree that $\mathbb{E}_{f,i}[d] = \theta_d$ and $\mathbb{E}_{f,i}[y_{f,i}] = 0$, as this is not our focus.[18] Rather, we focus on disagreement over expected inflation. Namely:

$$\mathbb{E}_{f,i}(\pi) = \mu_{f,i}\pi^e \quad \text{with} \quad \sum_{i\in\Theta_f} \mu_{f,i} + \sum_{i\in\Theta_b} \mu_{b,i} = 1 \tag{3}$$

such that the parameters $\mu_{f,i}$ capture this heterogeneity in inflation expectations and π^e is the *fundamental* expected inflation in this market. Note that this is without loss of generality, as it simply defines what π^e is as the average across expectations.

We assume that all the institutions believe that returns are normally distributed so that they solve a mean-variance optimization problem. The variances of the three exogenous random variables are σ_π^2, σ_d^2, and $\sigma_{y_{f,i}}^2$, while we denote the covariances of expected inflation with market returns and background risk by $\sigma_{\pi,d}$ and $\sigma_{d,y}$, respectively, and their associated correlations by $\rho_{\pi,d}$ and $\rho_{d,y}$.[19]

Finally, we allow for relatively general capacity constraints on each institution's ability to take on inflation risk. These capture regulatory constraints, balance sheet constraints, investment mandates, or limitations on short sales. The constraint is given by a continuous function:

$$G_f(q_{f,i}, z_{f,i}) \geq 0 \tag{4}$$

that measures the proximity of the pension fund to the capacity limits. The $z_{f,i}$ is an exogenous institution-specific shifter in the tightness of these financial constraints, while $q^*_{f,i}$ is the net notional exposure in equilibrium. We will use $\lambda_{f,i}$ to denote the Lagrange multiplier associated with this constraint at the optimal choice, and write $g_{f,i} \equiv \partial G_f(q^*_{f,i}, z_{f,i})/\partial q_{f,i}$ for brevity, omitting the arguments of the function.

A similar problem describes the actions of hedge funds in the short horizon market, where P, Π, and $Q_{h,i}$ are the prices, payoffs, and net holdings of short-horizon swaps, respectively, and $\Theta_h \in \mathbb{N}$ is the set of institutions in that sector. To be clear about the key restriction of *segmented markets* that we are imposing, and because we will refer to it later, the formal assumption is:

Assumption 1. *(Segmented markets.) Pension funds do not participate in the short-horizon market $Q_{f,i} = 0$ and hedge funds do not participate in the long horizon market $q_{h,i} = 0$.*

[18] Allowing these to be institution-specific would complicate the algebra, but leave the main message of all the propositions unchanged.

[19] Again to focus on inflation risk, and in order to reduce the length of the expressions, we set the correlation between market returns and background risk to zero: $\sigma_{d,y_{f,i}} = 0$.

Finally, for dealer banks, who are active in both the short and long horizon inflation markets, their budget constraint is:

$$a'_{b,i} = a_{b,i} + (\pi - p)q_{b,i} + (\Pi - P)Q_{b,i} + (d - s)e_{b,i} + y_{b,i} \tag{5}$$

so they choose both $q_{b,i}$ and $Q_{b,i}$. Following the large dealer net positions that we found in the data, especially in the long-horizon market, the dealers are solving a portfolio problem, as opposed to a market-making problem alone.

As with the client institutions, dealer banks also have capacity constraints on their ability to take on risk. However, given the high-frequency nature of our transaction-level data, we will assume that within each dealer bank, the ability of traders on the short horizon desk to take on risk is not constrained by the risk taken by the other desk serving the long horizon market within a day, prior to books being balanced. In reality, the desks at a dealer that sell long-horizon and short-horizon inflation swaps can be separated, and are often manned by different traders. Books are only compared at the end of day, and while orders are made to have the position in one book to be potentially offset or constraining the position in the other book, this happens only a day after. Formally, this implies that dealer banks face two separate capacity constraints.

Assumption 2. *(Desk separation within the day.) Dealers face separate capacity constraints:*

$$G_b^S(Q_{b,i}, z_{b,i}) \geq 0 \quad \textit{and} \quad G_b^L(q_{b,i}, z_{b,i}) \geq 0 \tag{6}$$

so that $\partial G_b^S(\cdot,\cdot)/\partial q_{b,i} = 0$ *and* $\partial G_b^L(\cdot,\cdot)/\partial Q_{b,i} = 0$.

3.2 Demand, supply and market clearing

Proposition 1, proven in Appendix C.1, states the solution of the optimization problem in the previous section in the form of an asset demand system of pension fund institutions.

Proposition 1. *Given market prices p^* and s^*, a pension fund's optimal demand for inflation protection $q^*_{f,i}$ scaled by size is given by:*

$$\frac{q^*_{f,i}}{a_{f,i}} = \underbrace{\frac{\mu_{f,i}\pi^e - p^*}{\gamma_{f,i}\sigma_\pi^2(1-\rho_{\pi,d}^2)}}_{\textit{price and beliefs}} - \underbrace{\left(\frac{\sigma_d}{\sigma_\pi}\right)\left[\frac{\theta_d - s^*}{\gamma_{f,i}\sigma_d^2(1-\rho_{\pi,d}^2)}\right]\rho_{\pi,d}}_{\textit{hedging demand}} - \underbrace{\left[\frac{1}{(1-\rho_{\pi,d}^2)\sigma_\pi^2}\right]\left(\frac{\sigma_{\pi,y_{f,i}}}{a_{f,i}} + \frac{\lambda_{f,i}g_{f,i}}{\gamma_{f,i}}\right)}_{\textit{liquidity frictions}} \tag{7}$$

Demand for inflation swaps scales with the size of the institutions, and depends on three terms.

The first is a subjective expected Sharpe ratio: the difference between expected inflation and the price of the swap, scaled by risk aversion times overall uncertainty. If an institution expects inflation to be higher it will want to buy more inflation protection; if it is more uncertain about inflation or more risk averse, it will respond less to those expectations. The slope of the demand curve in the traditional quantity-price space will be higher the more risk averse the agent is.

The second factor is the hedging of market risk. The higher the correlation between expected inflation and the returns of the pension fund's portfolio (higher $\rho_{\pi,d}$), the less it will want to buy inflation protection, since now higher inflation also comes with higher returns in other investments. This hedging demand scales with the size of the fund's position in the market, which depends on its Sharpe ratio.

The third factor captures liquidity frictions, driven by two features of the model. The first is that a higher covariance of expected inflation with background income lowers the demand for protection against inflation because this income gives a natural hedge. The second is that a binding capacity constraint lowers demand relative to what the pension fund would like to obtain but is prevented from doing so by regulations or internal governance constraints.

On the other side of the market are dealer banks. Their asset demand system looks similar, and is stated in Appendix C.2. Here, we present a slightly restricted version where their beliefs about how inflation at different horizons covaries with market returns follows a one-factor structure, so $\rho_{\pi,\Pi} = \rho_{\pi,d}\rho_{\Pi,d}$. Together with desk separation, this assumption leads to similar demand functions as those of the clients.

Proposition 2. *Given market prices P^*, p^* and s^*, the optimal allocation of risk in the two markets of a dealer bank is:*

$$\frac{q^*_{b,i}}{a_{b,i}} = \frac{\mu_{b,i}\pi^e - p^*}{\gamma_{b,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} - \frac{\sigma_d}{\sigma_\pi}\left[\frac{\theta_d - s^*}{\gamma_{b,i}\sigma^2_d(1-\rho^2_{\pi,d})}\right]\rho_{\pi,d} - \left[\frac{1}{(1-\rho^2_{\pi,d})\sigma^2_\pi}\right]\left(\frac{\sigma_{\pi,y_{b,i}}}{a_{b,i}} + \frac{\lambda^L_{b,i}g^L_{b,i}}{\gamma_{b,i}}\right) \tag{8}$$

$$\frac{Q^*_{b,i}}{a_{b,i}} = \frac{\mu_{b,i}\Pi^e - P^*}{\gamma_{b,i}\sigma^2_\Pi(1-\rho^2_{\Pi,d})} - \frac{\sigma_d}{\sigma_\Pi}\left[\frac{\theta_d - s^*}{\gamma_{b,i}\sigma^2_d(1-\rho^2_{\Pi,d})}\right]\rho_{\Pi,d} - \left[\frac{1}{(1-\rho^2_{\Pi,d})\sigma^2_\Pi}\right]\left(\frac{\sigma_{\Pi,y_{b,i}}}{a_{b,i}} + \frac{\lambda^S_{b,i}g^S_{b,i}}{\gamma_{b,i}}\right) \tag{9}$$

Asset prices in equilibrium are pinned down by market clearing conditions:

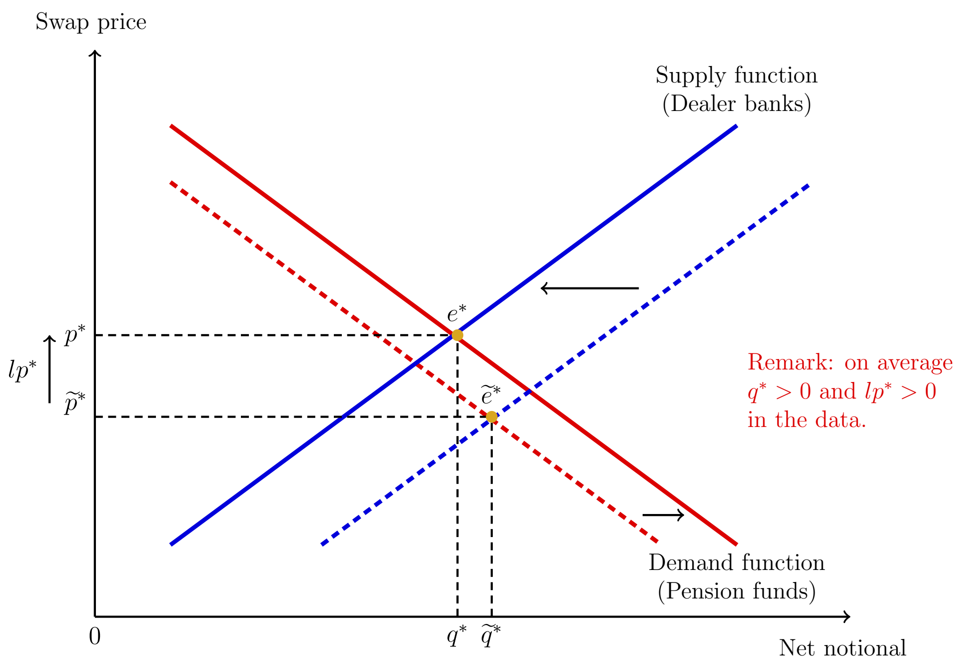
$$q^* \equiv \sum_{i\in\Theta_f} q^*_{f,i} = -\sum_{i\in\Theta_b} q^*_{b,i} > 0 \tag{10}$$

so that $q > 0$ means that pension funds have net positive notional holdings, and the negative of the sum of demand from dealer banks is the supply in the market.

3.3 The markets and the frictionless equilibrium

Figure 8 displays equilibrium in the long-horizon market with a simple graph that has the supply and demand curve as solid lines, and an equilibrium at the point e^*. In the data, we observed that $q^*_{b,i} < 0 < q^*_{f,i}$. The model can explain this as a result of four forces. First, pension funds may be more risk averse than banks, $\gamma_{f,i} > \gamma_{b,i}$. This is plausible, as they may be less well diversified and are forced by regulation to be more prudent. Second, pension funds have more hedging needs from their other assets than dealer banks, which makes sense given pension funds' large holdings of nominal bonds and the limited supply of index-linked bonds.[20] Third, pension funds are more exposed to background risk that covaries with inflation $\sigma_{\pi,y_{f,i}} > \sigma_{\pi,y_{b,i}}$, which again makes sense as many funds have liabilities denominated in real terms. Fourth, the regulatory environment for pension funds encourages them to buy inflation protection easing trading constraints: $\lambda_{f,i} < \lambda^L_{b,i}$. This is consistent with real world pension fund practices, such as liability-driven investments. Overall, therefore, the model can make sense of what we see in the data.

Figure 8 THE FRICTIONLESS EQUILIBRIUM



The *frictionless market equilibrium* arises when there are complete markets to fully insure institution-specific income risk, so $\sigma_{\pi,y_{b,i}} = \sigma_{\pi,y_{f,i}} = \sigma_{\pi,y_{h,i}} = 0$, and the regulatory, short-sale or other capacity constraints do not bind for any agent in either the long horizon market, $\lambda^L_{b,i} = \lambda_{f,i} = 0$, or the short horizon market, $\lambda^S_{b,i} = \lambda_{h,i} = 0$. Appendix C.3 solves for this counterfactual.

[20]For instance, see `https://prod.schroders.com/en/sysglobalassets/schroders/sites/ukpensions/pdfs/2016-06-pension-schemes-and-index-linked-gilts.pdf`

Lemma 1. *If $\widetilde{p}$ is the frictionless price of a long horizon inflation swap, in equilibrium, it is:*

$$\widetilde{p}^* = \underbrace{\left[\frac{\sum_{i\in\Theta_f}\widetilde{\gamma}_{f,i}^{-1}\mu_{f,i}}{\sum_{i\in\Theta_f}\widetilde{\gamma}_{f,i}^{-1}+\sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}} + \frac{\sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}\mu_{b,i}}{\sum_{i\in\Theta_f}\widetilde{\gamma}_{f,i}^{-1}+\sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}}\right]}_{\equiv\Lambda\ \text{,size-weighted dispersion of beliefs}} \underbrace{\pi^e}_{\text{expected inflation}} - \underbrace{\frac{\theta_d-\widetilde{s}^*}{\sigma_d^2}\sigma_{\pi,d}}_{\text{risk premium}} \tag{11}$$

The frictionless price for swaps depends on expected inflation minus a compensation for risk. Importantly, the coefficient on π^e is a weighted function of each institution's beliefs about expected inflation that depend in turn on the size of their gross notional risk traded. Therefore, the first term on the right hand side does not equal π^e necessarily, as long as there is granularity in asset holdings with some institutions being much larger, so that their beliefs carry much more weight in pricing of the asset. We use ε_π to denote the fundamental shocks that drive the frictionless price.

Given this frictionless price, the liquidity premium lp^* is then the difference between it and the actual price:

Lemma 2. *The liquidity premium in the long horizon market is defined as $lp^* = p^* - \widetilde{p}^*$, and is driven by frictions in both the demand from pension funds and the supply from dealer banks:*

$$lp^* = \underbrace{-\frac{\sum_{i\in\Theta_b}\left\{\sigma_{\pi,y_{b,i}}+\frac{\lambda_{b,i}^L g_{b,i}^L}{\widetilde{\gamma}_{b,i}}\right\}}{\sum_{i\in\Theta_f}\widetilde{\gamma}_{f,i}^{-1}+\sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}}}_{\equiv\varepsilon_b,\ \text{the supply friction from dealer banks}} + \underbrace{-\frac{\sum_{i\in\Theta_f}\left\{\sigma_{\pi,y_{f,i}}+\frac{\lambda_{f,i} g_{f,i}}{\widetilde{\gamma}_{f,i}}\right\}}{\sum_{i\in\Theta_f}\widetilde{\gamma}_{f,i}^{-1}+\sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}}}_{\equiv\varepsilon_f,\ \text{the demand friction from pension funds}} \tag{12}$$

Figure 8 represents the frictionless supply and demand as dashed lines, and the equilibrium as $\widetilde{e}^*$. The literature has emphasized the capacity constraint facing dealers, in which case $\lambda_{b,i}^L g_{b,i}^L$ would be the dominant term, and the liquidity premium is positive.[21] This is the case depicted in the figure, where the actual supply is further to the left of the frictionless one than is the case with demand ($\varepsilon_b > \varepsilon_f$). However, shocks that are specific to the pension fund sector may sometimes make its capacity constraint bind sharply, $\lambda_{f,i}g_{f,i}$ jumps, and the liquidity premium can easily turn negative (ε_f dominates). Moreover, changes in the macroeconomy, in monetary policy, and in the policy regime driving inflation, which have been plentiful during our sample period, will change the perceived correlation between inflation and the income flows of pension funds and banks, so that $\sigma_{\pi,y_{f,i}}$ and $\sigma_{\pi,y_{b,i}}$ can be large and volatile, leading to large fluctuations in the liquidity premium. The observed prices of inflation swaps p^* can therefore be very far from actual risk-adjusted expected inflation $\widetilde{p}^*$, and move significantly over time, driven by market frictions and institutional-level constraints that shift both demand and supply.

[21]This corresponds to a *negative* supply shock.

3.4 The response of prices and quantities to shocks

The same ingredients drive the short-horizon inflation market where hedge funds and dealer banks meet and markets clear according to the condition:

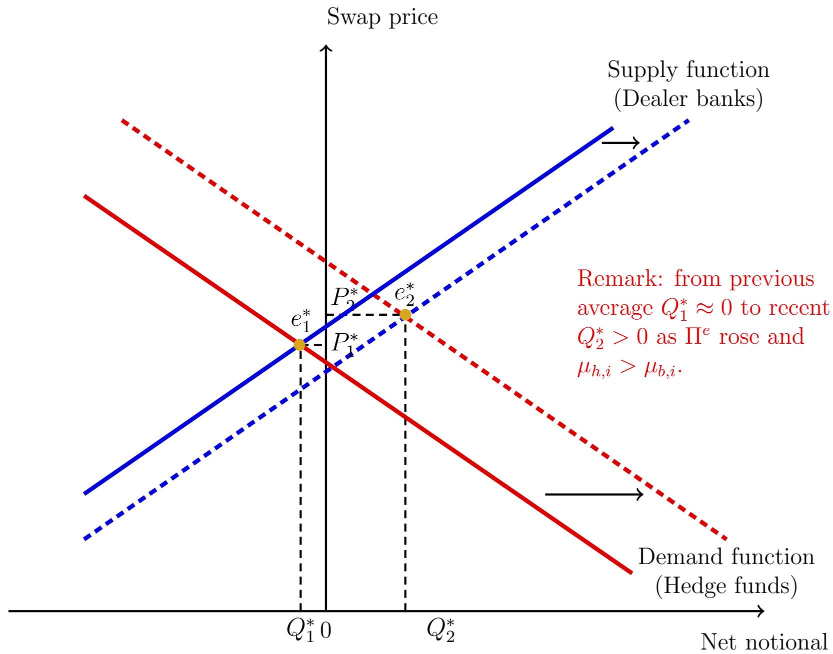
$$Q^* = \sum_{i \in \Theta_h} Q^*_{h,i} = -\sum_{i \in \Theta_b} Q^*_{b,i} \tag{13}$$

The equilibrium is represented in Figure 9, starting from an initial situation with solid lines and subscript 1. Banks and hedge funds have on average similar beliefs regarding hedging demand, background risk, regulation. Therefore, the model would predict that equilibrium net holdings Q^*_1 are on average close to zero, as is the case in the data.

After a fundamental shock, demand and supply shift to the shaded lines, and the new equilibrium has a subscript 2. We can expect this market to have large swings, in quantities and prices, because changes in expected inflation π^e come with large changes in the dispersion of beliefs. This is because hedge funds, being more informed traders, would be expected to be more sensitive to inflation news than banks: $\mu_{h,i} > \mu_{b,i}$, matching the recent shift to a positive Q^*_2.

At the same time, the pandemic and political uncertainty have shifted the liquidity frictions for both sides of the market. It is conceivable that liquidity frictions have loosed for both, and relatively more so for hedge funds, matching the shift in demand and supply in the figure just as well as if there had been a fundamental shock. Therefore, observing the new price P^*_2, there is a difficult identification problem in measuring how much has expected inflation actually increased.

Figure 9 Equilibrium in the short horizon inflation swap market



4 Three identification strategies

The identification problem is that the observed prices of inflation swaps p and P can move because of shocks to fundamentals or shocks to the liquidity frictions. As the model clarified, there are three separate sources of frictions, from each of the three sectors trading in the market. There are four shocks in column vector of shocks is: $\boldsymbol{\varepsilon} = (\varepsilon_h, \varepsilon_f, \varepsilon_b, \varepsilon_\pi)'$ and usually only two observables.

We, instead, have daily data on prices *and* quantities from the 2^{nd} January 2019 to the 10^{th} February 2023, for 1078 observations. Therefore, unlike previous studies, we have four variables corresponding to the following aggregated data series: (i) q: net purchases of UK RPI inflation swaps by PFLDIs with initial time-to-maturity 10 years or more; (ii) p: a weighted-average daily price of 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 PFLDI institutions as a share of the total across the data sample; (iii) Q: the net purchases of swaps by hedge funds with initial time-to-maturity 3 years or less; and (iv) P, 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. Swap prices corresponding to inflation swaps of various maturities have also been adjusted for the indexation lags that follow from the market convention for RPI fixings, so that they are purely a forward-looking measure of RPI inflation expectations.[22] Let $\mathbf{Y} = (Q, P, q, p)'$ be the column vector with these variables.

[22] We thank Carolin Pflueger for this suggestion. More details regarding the market convention of UK RPI swap pricing and our adjustment methodology are presented in Appendix A.2.

Having as many variables as shocks is progress, but still not enough for identification. The fundamental shocks shift all demand and supply curves, while the liquidity shocks shift the demand for either pension funds, hedge funds, or dealers, separately depending on their source. Formally:

$$\mathbf{Y} = \mathbf{\Psi} \boldsymbol{\varepsilon} \tag{14}$$

and we need to pin down the elements of the 4×4 matrix $\mathbf{\Psi}$ to fully identify the system, or at least four elements in its inverse to extract from the data $\mathbf{Y}$ the fundamental expected inflation ε_π. In this section we describe how to use the high-frequency, the cross-section, and the time length of our data to achieve identification through three separate strategies that exploit each of these three features.

4.1 First identification strategy: heterogeneity in reactivity

It seems plausible to believe that in response to fundamental news, dealers are more informed (or attentive, or reactive) than pension funds. After all, the dealers trade the inflation risk and see all sides of the market. As such, they have more precise posterior information about inflation and they respond more to news. In this case, following a shock to the fundamental that raises π^e (and so frictionless $\widetilde{p}$), the upward shift in the supply function dominates the shift in the demand function. Therefore, we expect the new equilibrium to have a higher price and a lower net notional amount traded. That is, p should rise and q should fall.

Instead, in the short horizon market, the informed hedge funds are likely more reactive to fundamentals than dealer banks, as we discussed in Figure 9. Therefore, now it is the upward shift in the demand function that dominates leading to a higher P and a rise in Q.

This identifies a fundamental shock as the one that satisfies these sign restrictions on quantities following a rise in prices (and symmetrically for a fall). Formally, the identification assumption is:

Assumption 3. *(Differential reactiveness to fundamental news about inflation.) Dealer banks respond more to fundamental long-horizon expected inflation than pension funds but less to fundamental short-horizon expected inflation than hedge funds:*

$$\frac{\sum_{i\in\Theta_h} \widetilde{\gamma}_{h,i}^{-1} \mu_{h,i}}{\sum_{i\in\Theta_h} \widetilde{\gamma}_{h,i}^{-1} + \sum_{i\in\Theta_b} \widetilde{\gamma}_{b,i}^{-1}} > \frac{\sum_{i\in\Theta_b} \widetilde{\gamma}_{b,i}^{-1} \mu_{b,i}}{\sum_{i\in\Theta_f} \widetilde{\gamma}_{f,i}^{-1} + \sum_{i\in\Theta_b} \widetilde{\gamma}_{b,i}^{-1}} > \frac{\sum_{i\in\Theta_f} \widetilde{\gamma}_{f,i}^{-1} \mu_{f,i}}{\sum_{i\in\Theta_f} \widetilde{\gamma}_{f,i}^{-1} + \sum_{i\in\Theta_b} \widetilde{\gamma}_{b,i}^{-1}} \tag{15}$$

The other two assumptions, already discussed, identify the other shocks in the model. Suppose that there is a shock to the liquidity component of demand from pension funds that leads to a rightward shift of the demand function. In equilibrium, this would raise p and q simultaneously in the long horizon market. Therefore, for the same increase in p, we can use the equilibrium responses in quantities q to distinguish between a liquidity demand shock and a fundamental shock. If instead the shock were to the liquidity component of dealers' supply of inflation protection, this would shift the supply curve downward, and so lead to a higher p and lower q, just as with a fundamental shock. To separately identify the two, we use assumption 1 on segmented markets together with the price and quantity information from the short-horizon market. Since the shock to the dealers would *also* appear in the short-horizon market, we should observe P rising and Q falling.[23] Observing both markets, we can therefore distinguish between the two shocks.

There is a final barrier to overcome for identification. If the change in the supply of inflation swaps in the long market following a fundamental shock causes the capacity constraints in the short market to bind, this would lead to a reduction in supply in the short horizon market. Therefore, P would rise and Q fall, making the supply shocks indistinguishable from a fundamental shock. This is ruled out by the desk-separation assumption 2: within any given trading day, this spillover from quantity supplied in one market into constraints in the other market does not happen.

Note that all three assumptions exploit the high frequency in the data. We should definitely not expect that desks are separated within a month, or even a week. Also, whatever information differences between banks, hedge funds, and pension funds may well be gone within a few days; in fact, we will find it is so. Combining the discussion of demand and supply shifters in the model, the three assumptions imply the following sign restrictions on the response of observables to shocks within a day:

$$\mathbf{\Psi} = \begin{pmatrix} + & 0 & - & + \\ + & 0 & + & + \\ 0 & + & - & - \\ 0 & + & + & + \end{pmatrix} \tag{16}$$

These sign restrictions set identify the four shocks, and the response of each variable to them.

4.2 Second identification strategy: cross-sectional granularity

An alternative way to go about identifying the fundamental shocks uses the buying and selling behavior of specific market participants. If the previous strategy exploited the high frequency of the data, this strategy exploits its cross-sectional richness. Assumption 3 on differential reactiveness within a day is replaced by a new assumption on cross-sectional granularity.

Before we state this assumption, we first re-write the asset demand system in proposition 1 as a reduced-form factor model. Let $\varepsilon_{f,i,t}$ be pension fund i's demand shock arising from liquidity

[23]Formally, recall that the shock $z_{b,i}$ to dealer bank i's ability to take on inflation risk affects the capacity constraints in both markets: $G_b^S(Q_{b,i}, z_{b,i})$ and $G_b^S(Q_{b,i}, z_{b,i})$.

frictions, defined as:

$$\varepsilon_{f,i,t} = -\frac{1}{(1-\rho_{\pi,d}^2)\sigma_\pi^2}\left(\frac{\sigma_{\pi,y_{f,i}}}{a_{f,i,t}} + \frac{\lambda_{f,i}g_{f,i,t}}{\gamma_{f,i}}\right) = \frac{\kappa_{f,i}^{lp} lp_t^*}{\gamma_{f,i}\sigma_\pi^2(1-\rho_{\pi,d}^2)} + \tilde{\varepsilon}_{f,i,t} \tag{17}$$

The second equality separates this demand shock into the sum of an *idiosyncratic* disturbance $\tilde{\varepsilon}_{f,i,t}$ and a market-wide liquidity premium lp_t^* (defined in Equation (12)) with a fund-specific impact $\kappa_{f,i}^{lp}/\gamma_{f,i}$. With these definitions, the demand system in Equation (7), appended with time subscripts, can be rewritten as:

$$\frac{q_{f,i,t}}{a_{f,i,t}} = \boldsymbol{\omega}_{f,i}'\mathbf{F}_t + \tilde{\varepsilon}_{f,i,t} \tag{18}$$

The $\mathbf{F}_t$ are unobserved common factors and the $\boldsymbol{\omega}_{f,i}$ are the fund-specific factor loadings, defined by:

$$\mathbf{F}_t = \begin{pmatrix} \pi_t^e \\ lp_t^* \end{pmatrix}, \quad \boldsymbol{\omega}_{f,i} = \begin{pmatrix} \frac{(\mu_{f,i}-\Lambda)}{\gamma_{f,i}\sigma_\pi^2(1-\rho_{\pi,d}^2)} \\ \frac{\kappa_{f,i}^{lp}-1}{\gamma_{f,i}\sigma_\pi^2(1-\rho_{\pi,d}^2)} \end{pmatrix} \tag{19}$$

In the implementation, we estimate this regression with an interactive fixed effects model following Bai (2009) to extract the idiosyncratic component of *each* pension fund's demand. We subsequently construct an instrument for sector-wide shocks to pension fund demand $\varepsilon_{f,t}$ defined in equation (12), as a weighted sum of the residuals:

$$GIV_{f,t} = \sum_{i\in\Theta_f} a_{f,i,t}\tilde{\varepsilon}_{f,i,t} \tag{20}$$

Given that $\mathbf{F}_t$ spans all the determinants of demand by the pension funds, we know that the residuals are independent of fundamentals $\mathbb{E}(GIV_{f,t}\varepsilon_{\pi,t}) = 0$ and of the determinants of the supply by banks: $\mathbb{E}(GIV_{f,t}\varepsilon_{b,t}) = 0$. Further, because of the segmentation of markets (assumption 1), we know that $\mathbb{E}(GIV_{f,t}\varepsilon_{h,t}) = 0$. Therefore, the exclusion restriction for $GIV_{f,t}$ to be an instrument for $\varepsilon_{f,t}$ is satisfied.

For the instrument to be relevant though requires $GIV_{f,t}$ to be correlated with $\varepsilon_{f,t}$. Yet, from the properties of estimated regression residuals, we know that: $\sum_{i\in\Theta_f}\tilde{\varepsilon}_{f,i,t} = 0$. If the data is granular in the sense that the the average of these residuals is no longer zero once they are weighted by the size of each firm, then the relevance condition can be satisfied. This is because the shocks to some of the large firms can drive the average. The same applies to $GIV_{h,t}$ and $GIV_{b,t}$ for the liquidity shocks to hedge funds and banks, respectively. We then have three valid instruments, which by the properties of a system that has four variables and four shocks, implies that the shocks to fundamentals are point identified.

The intuition behind this approach is that because some institutions are larger, individual shocks to their demand function will drive the aggregate demand in the market. Since we have many such institutions, we can measure the idiosyncratic changes in their demand. This provides an aggregate demand shock. This is possible because we have institution-specific transaction data on $q_{f,i,t}$ and $a_{f,i,t}$ and so we can identify these large institutions.[24]

More precisely, the assumption of granularity is:

Assumption 4. *(Granularity of the institutions.) The data on asset positions $a_{f,i,t}$, $a_{h,i,t}$ and $a_{b,i,t}$ is granular in that:*

$$\mathbb{E}(GIV_{f,t}\varepsilon_{f,t}) \neq 0 \ \ \textit{and} \ \ \mathbb{E}(GIV_{b,t}\varepsilon_{b,t}) \neq 0 \quad \textit{and} \ \ \mathbb{E}(GIV_{h,t}\varepsilon_{h,t}) \neq 0 \tag{21}$$

We can verify the plausibility of this assumption in the data. In our sample, there are 210 PFLDIs. Figure 10 shows the plot of the rank of each PFLDI institution against their outstanding gross notional positions. The Pareto parameter on their outstanding gross notional position is 0.13, strongly supporting the granularity assumption. For institutions within the PFLDI sector with an outstanding gross notional position to-date larger than \$1bn, we estimated a power law regression of (the log of) their rank on (the log of) their gross notional positions (Gabaix, 2016). The fit of the regression is also in the figure. The estimated power law coefficient is -0.9, with a standard error of 0.013 and an R^2 of 0.979. Therefore, the size of PFLDI's gross inflation risk exposures comes close to satisfying Zipf's law, which is a particular power law distribution with a power law coefficient of -1.

There are fewer hedge funds (30) and banks (16), so results are more imprecise. Still, the power law exponent points estimates for hedge funds and dealer banks are -0.728 and -0.402 with standard errors 0.035 and 0.058, respectively, again supporting granularity.

An alternative approach to check the relevance of the instruments is through standard first-stage F-statistics. They are: 72.3 for $GIV_{f,t}$, 22.3 for $GIV_{h,t}$ and 43.5 for $GIV_{b,t}$.

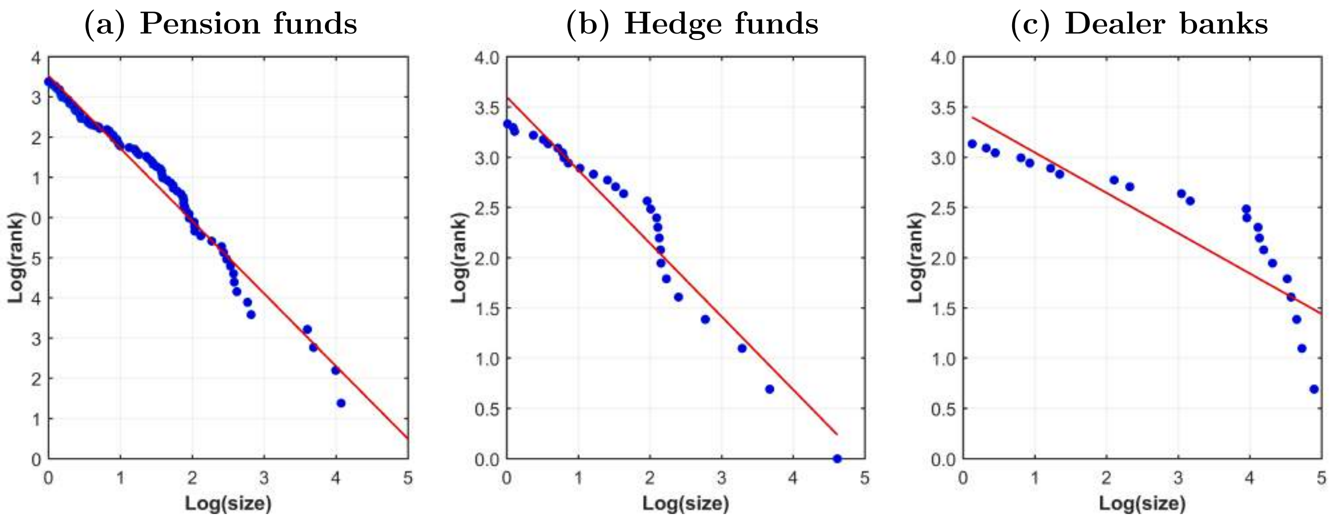
4.3 Third identification strategy: heteroskedasticity across time

The third identification strategy exploits instead the time series of 1078 days. The new assumption, to complement market segmentation and desk separation, is that fundamental news about inflation are lumpy. The underlying shocks may themselves be smooth over time, but it is around data release dates that traders learn the most about where inflation is heading.

In the data, at these dates, the volatility of inflation swap prices is noticeably higher. While liquidity shocks may also be higher during those dates, the major difference should be that actual

[24] A more technical discussion of the identification strategy and details of its implementation is in Appendix D.

Figure 10 Institutional rank versus outstanding gross notional positions



Note: Size refers to the gross notional position outstanding (in USD bn) that the institution has acquired in the UK RPI market up to the latest date of our data sample. Each scatter marker refers a given institution, and the line in red denotes the fitted value. To arrive at each institution's gross outstanding position in the inflation swap market, we first construct an unbalanced panel by tracking the trading activity of each institution across various execution dates of the trade, and then cumulatively construct a stock of their outstanding positions while taking into consideration older trades that expire. Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

fundamentals are revealed on these dates. The identifying assumption is that this spike in variance is driven by the emergence of fundamental shocks on these dates, which again seems reasonable. This allows for partial identification of the fundamental shocks alone. More formally:

Assumption 5. *(Heteroskedascity at known dates due to fundamentals.) If $\mathbf{\Sigma}_h$ is the variance-covariance matrix of the shocks $\boldsymbol{\varepsilon}$ at data release dates, and $\mathbf{\Sigma}_l$ the one at other dates, then the largest diagonal element of $\mathbf{\Sigma}_h'\mathbf{\Sigma}_l$ is the one associated with the variance of the fundamentals ε_π.*

In our sample, we have 48 monthly dates when the data on UK RPI inflation was released. As a first check, we indeed observe spikes in trading activity along with large price adjustments in the vicinity of these dates. In addition to these 48 dates, one more date in our sample had a large impact: 6^{th} September 2022 when ex-Prime Minister Truss announced a cap on energy prices that dramatically changed the properties of measured inflation.

To check the plausibility of assumption 5, we estimate a vector autoregression on those 49 dates, as well as a second VAR on the remaining dates. Calculating the product of the inverse of the variance-covariance matrix of residuals on the latter dates and the variance-covariance matrix of residuals on the former dates, the resulting matrix has a largest eigenvalue of 4.651. That is, there is clearly more variation in both prices and quantities during RPI release dates. As a robustness check, we also include the 8 dates from each year of our sample when the Bank of England's Monetary Policy Committee (MPC) announces their decisions for monetary policy. Figure F.5 in Appendix F.2 shows that our estimation results are nearly unchanged when these MPC announcement dates are additionally included to the set of RPI release dates. This further highlights that it is the heteroskedasticity from RPI release dates that leads to identification of the fundamental shock.

Finally, following Lütkepohl, Meitz, Netšunajev, and Saikkonen (2021), we implement a Wald-type test under the null hypothesis of homoskedasticity and so no identification. The null is rejected at a 0.1% significance level.[25]

4.4 Dynamics and implementation

So far, we have discussed the model and identification in a purely static setting. If markets were efficient after one day, then this would suffice. However, perhaps information diffuses more slowly over several days, and capacity constraints and regulations may involve multiple successive days. The empirical model must be dynamic to allow for these features.

We estimate a Bayesian SVAR with diffuse priors, a deterministic constant $\mathbf{c}$, and $L = 3$ lags (picked by a Bayesian information criterion):

$$\mathbf{Y}_t = \mathbf{c} + \sum_{\ell=1}^{L} \mathbf{\Phi}_\ell \mathbf{Y}_{t-\ell} + \mathbf{u}_t \quad \text{and} \quad \mathbf{u}_t = \mathbf{\Psi}\boldsymbol{\varepsilon}_t. \tag{22}$$

Then, we implement the timing identification following Arias, Rubio-Ramírez, and Waggoner (2018), where sign restrictions on $\mathbf{\Psi}$ give set identification. For the granularity identification, we follow Stock and Watson (2018) and use the GIV as proxy instrumental variables. For the heteroskedasticity identification, we follow Brunnermeier, Palia, Sastry, and Sims (2021) and pick the shock that most closely satisfies the properties of a fundamental shock in our model.

To be able to compare the estimation results across the three identification strategies, we scale the impulse response functions in the following manner. The increase in the net notional position aggregated across all client institutions in the short horizon inflation market is always scaled to equal \$1bn in response to either the fundamental shock, the demand shock from hedge funds, or the supply shock from dealers. As for the demand shock from PFLDIs in the long horizon market, it is scaled such that it raises the net notional position aggregated across all client institutions in the long horizon inflation market by \$1bn.

[25] The corresponding test statistic is 230.9 when using only RPI release dates, and 108.9 when using both MPC announcement and RPI release dates. Given so, we use only RPI release dates as the benchmark specification for our heteroskedastictiy-based identification.

4.5 Cross-verification of the identification assumptions

The promise of having three identification strategies is not only that they can challenge the robustness of the findings, but also that they can internally cross-verify the identification assumptions that each strategy has independently imposed. Given any one of the identification strategies, we can let the data speak on whether the identification assumptions made by the other strategies seem to hold.

Starting with assessing the identification based on assuming differential reactiveness to fundamental shocks, the impulse response functions to a fundamental shock estimated under the other two identification strategies both have a rise in the net notional traded in the short-horizon market and a reduction in this quantity in the long-horizon market. Prices rise on impact in both markets, and the confidence bands are such that we do not reject the sign restrictions imposed by differential reactiveness on the estimates from the other two identification strategies.[26]

Next, turning to the identification based on the heteroskedasticity assumption, we find evidence that the fundamental shocks identified by sign restriction are more volatile on the RPI release dates. Across the 10,024 importance sampler draws that give set identification, the ratio of the variance of the fundamental shock on these dates to that obtained from other dates remaining in our data sample has a median of 1.346, and is above one 98.75% of the times. For several draws, the variance of the fundamental shock can be nearly *three* times as large when compared to the non-inflation release and news dates (see Appendix E.11).

Third, we investigate the granularity assumption using the fundamental shocks $\hat{\varepsilon}_{\pi,t}$ estimated from the other two approaches. We compute the sample analogs of the exclusion restriction conditions by checking whether:

$$\frac{1}{T}\sum_{t=1}^{T} GIV_{\nu,t}\hat{\varepsilon}_{\pi,t} = 0, \quad \text{for} \quad \nu \in \{f, h, b\} \tag{23}$$

Using the draws of the estimated fundamental shocks from set identification, the median estimates are -0.0092, 0.0263, and 0.0110 for $\nu = f$, h, and b, respectively.[27] Using instead the point estimates of the fundamental shock identified via heteroskedasticity, these are 0.0382, -0.1079 and -0.0384, respectively. All are close to zero, supporting the validity of the granular instrumental variables.

Finally, we conclude this section by reporting the absolute value (since shocks are identified up to a sign) of the correlation of the estimated fundamental shocks obtained from each of the

[26] These impulse responses are reported in Figure 11.

[27] Figure E.13 in Appendix E.11 shows the posterior distributions of the sample analogs computed across all draws.

three strategies.[28] Ordering the fundamentals from sign restriction strategy first, GIV strategy next, and heteroskedasticity strategy last, these correlations are:

$$\begin{bmatrix} 1 & 0.9865 & 0.8038 \\ \cdot & 1 & 0.7320 \\ \cdot & \cdot & 1 \end{bmatrix} \tag{24}$$

Such a remarkably high degree of correlation across three estimates that come from completely different identification strategies, using completely different properties of the data, give confidence that the estimates that follow are robust.

5 Estimates: the structure of the inflation swap markets

This section presents our estimates and findings on the main features of the UK RPI inflation swap market.

5.1 Impulse responses: the speed of adjustment to fundamentals

Figure 11 shows the impulse response to a fundamental shock for the three identification strategies. They are all qualitatively similar. Under our timing identification strategy, recall that the sign restricted responses of prices in both markets should increase, while those of the quantities traded in the long and short horizon markets should be negative and positive, respectively. This is what the differential reactiveness assumption implies, and, as we noted earlier, the signs of the responses are verified by the other identification restrictions.

There are significant differences in the size of the price responses across strategies. They are larger when identified with the granular strategy, and smaller with the heteroskedasticity strategy, with the estimates from the timing strategy in between. However, the error bands are large and intersect across the three sets of price impulse responses, at both the short and long horizons. Because this pattern of very similar qualitative responses is true across other experiments, from now onwards we report only results with the timing identification strategy to conserve space. All the conclusions that we draw are robust to the identification strategy.

Perhaps the most striking result from these impulse response is that the market completely adjusts quickly. All responses stabilize within one to three days and then persist over time. This suggests that fundamental shocks are permanent and that information diffuses quite quickly. Given the high frequency of the data, perhaps the persistence is not too surprising: after news

[28] The actual identified shocks are shown in Appendix E.10.

about inflation arrives, often in a lumpy way at a data release date, agents will persistently adjust their expectations of inflation until at least the next data release the following month. The quick adjustment of prices is more surprising as it suggests that this market is close to being (weakly) informationally efficient.

5.2 Impulse responses: liquidity shocks and the slopes of the demand and supply functions

Figure 12 shows the estimated dynamic responses to the demand and supply liquidity shocks. These conform with the standard responses one would expect from shocks to supply and demand. Namely, swap prices rise and the quantity traded falls in both markets following a supply shock originating from dealers that would shift the supply function upwards, while both rise in response to a demand shock.[29]

The estimated impulse response functions to a liquidity shock to dealers locally identifies the slope of the demand curves. The intuition is as follows. A liquidity shock to dealers is a reduction in supply, so it will trace out the demand functions in both the long and short horizon markets on impact. In the long horizon market, we can compute the slope of the demand function locally by calculating dp/dq, where dp and dq are the impact responses of the swap breakeven rate and aggregate quantity traded. The same applies to the slope of the demand function in the short horizon market given by dP/dQ, which we get from the relative responses of prices and quantities in the short horizon market. The same applies to the supply function in the long horizon and short horizon market, using the liquidity shocks to pension funds and hedge funds, respectively.

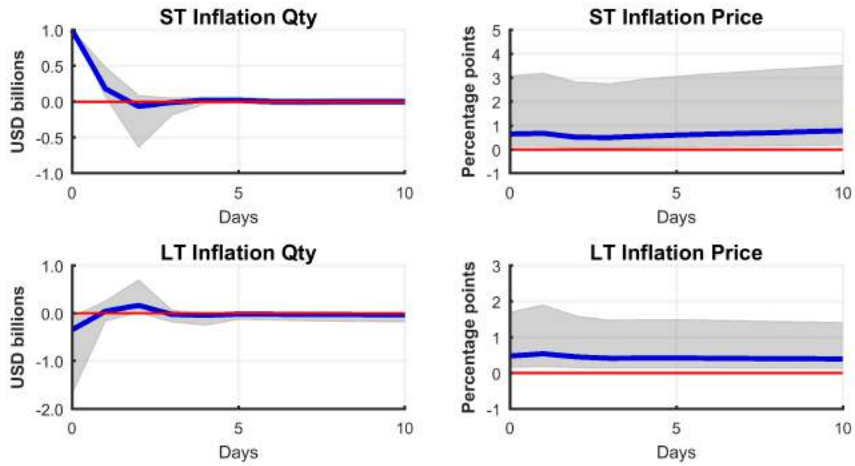
Figure 13 shows the estimated posterior distributions of these slopes from the Bayesian SVAR. In the short-horizon market, dealer banks and hedge funds are similarly price-sensitive. Perhaps this is a reflection of sellers and buyers in this market being similarly sophisticated and having similar bargaining power when negotiating in this over-the-counter market. Comparing across markets, the slope of the demand function in the short horizon markets by hedge funds is only slightly higher than the slope of the demand function in the long horizon market by pension funds.

The more striking result comes instead from the slope of the supply function in the long horizon market: it is close to zero, and much lower than the corresponding slope in the short horizon market. What drives this sharp result in the data? When we identify liquidity shocks to pensions funds, for instance by seeing large idiosyncratic changes in the holdings of the largest pension funds in our granularity strategy, these come with large changes in the net holdings but almost no change in the inflation swap prices.

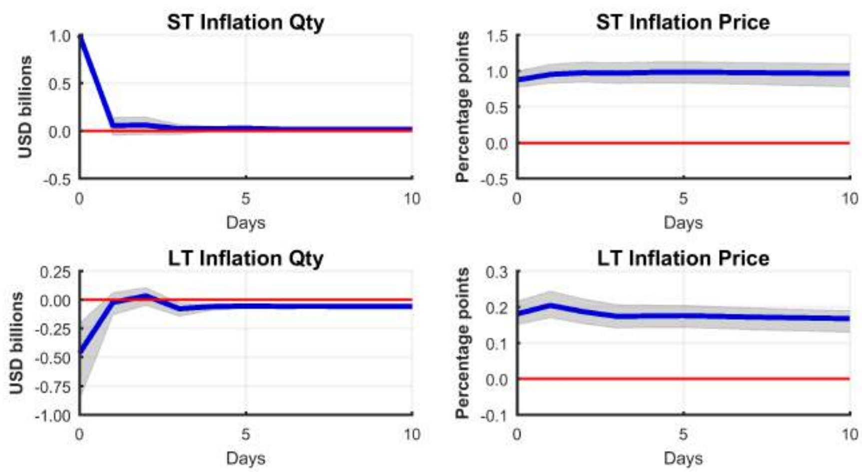
[29]Note that, on impact, we have left the responses of price and quantity from the other market unrestricted in the timing restriction strategy. Hence, this is a result, not an assumption.

Figure 11 ESTIMATED IMPULSE RESPONSE FUNCTIONS TO A FUNDAMENTAL SHOCK

(a) With timing restriction strategy



(b) With heteroskedasticity strategy



(c) With granular identification strategy

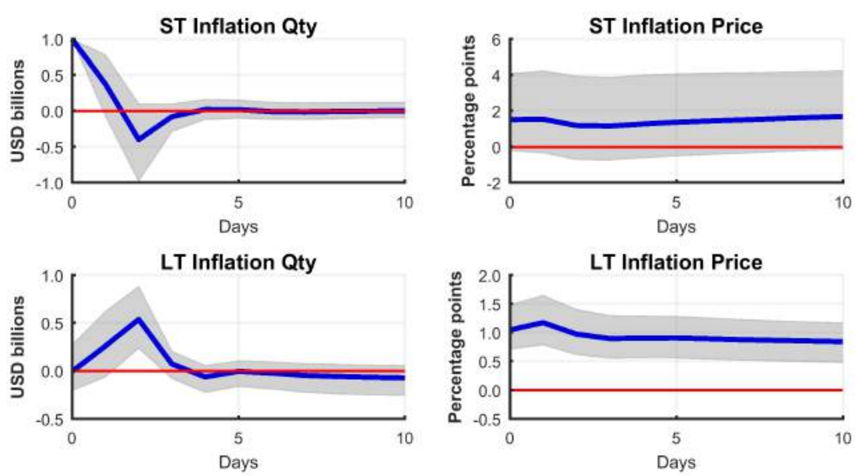
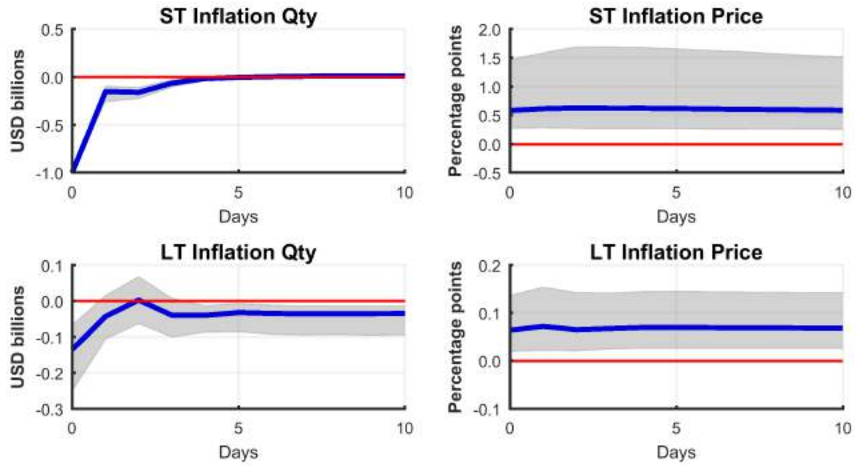
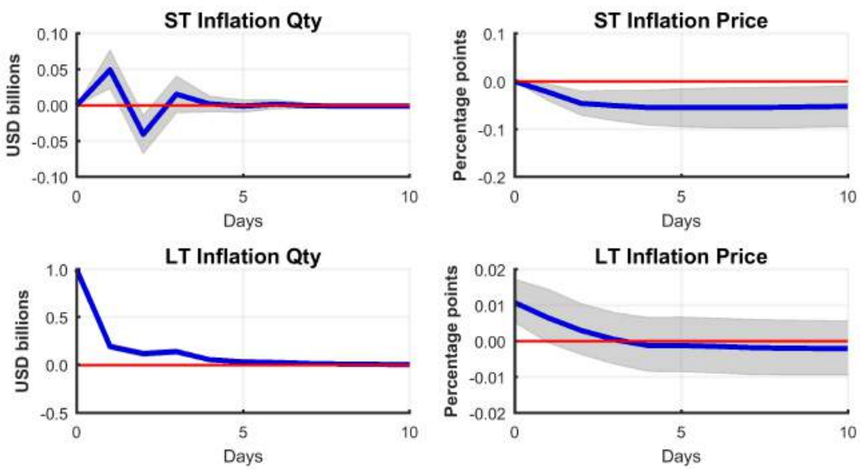


Figure 12 ESTIMATED IMPULSE RESPONSE FUNCTIONS TO LIQUIDITY SHOCKS

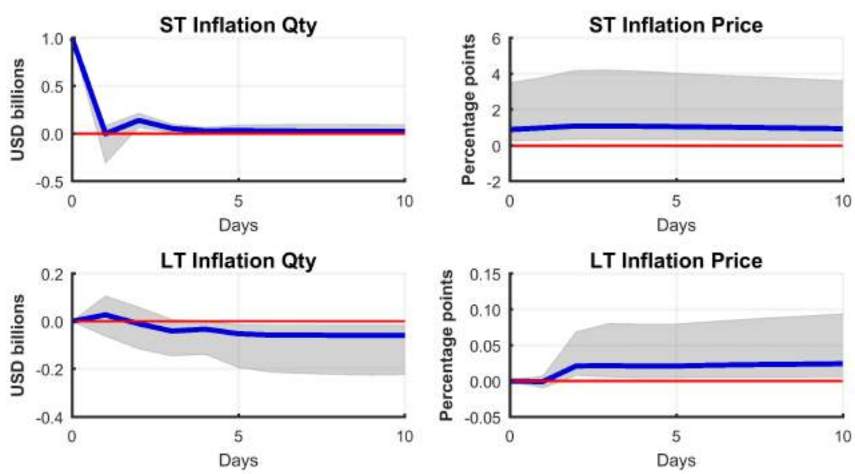
(a) Dealer supply shock ($\varepsilon_{b,t}$)



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



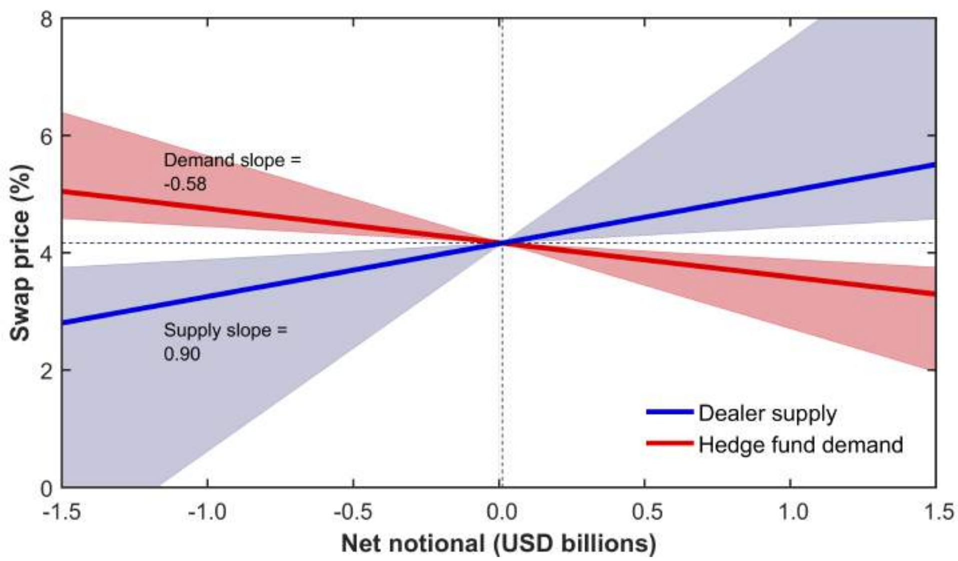
(c) Hedge fund demand shock ($\varepsilon_{h,t}$)



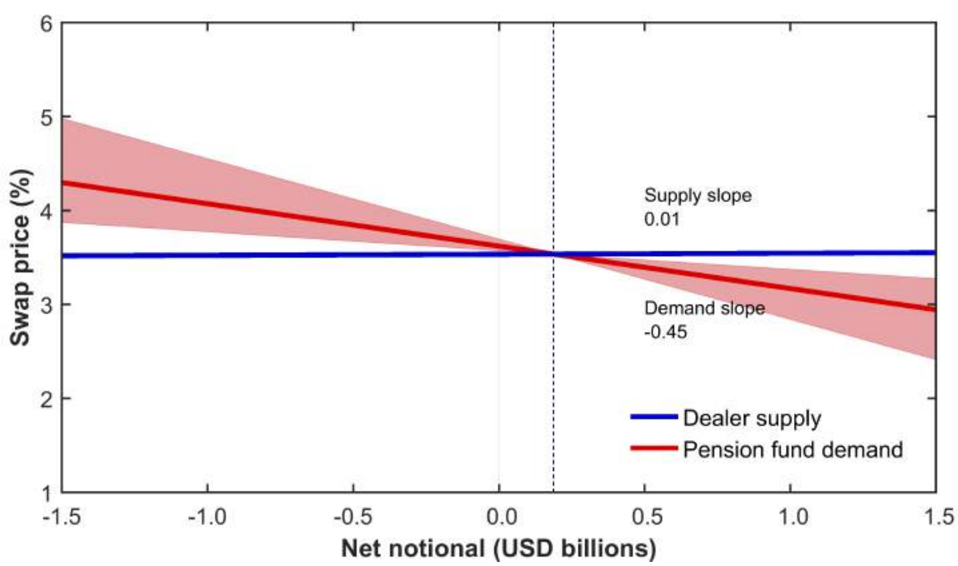
One interpretation of what lies behind this almost-horizontal supply curve is that dealers effectively set prices in the long-horizon market, with full bargaining power relative to their pension fund clients. Another interpretation, along the lines of preferred-habitat models, is that the risk capacity of arbitrageurs is larger with respect to long-horizon inflation risk than short-horizon risk. Making more progress in explaining this striking new facts requires a more detailed model of the industrial organization of these markets. We leave this exciting challenge for future research.

Figure 13 Estimated slopes of market demand and supply functions

(a) Short horizon UK RPI inflation swap market



(b) Long horizon UK RPI inflation swap market

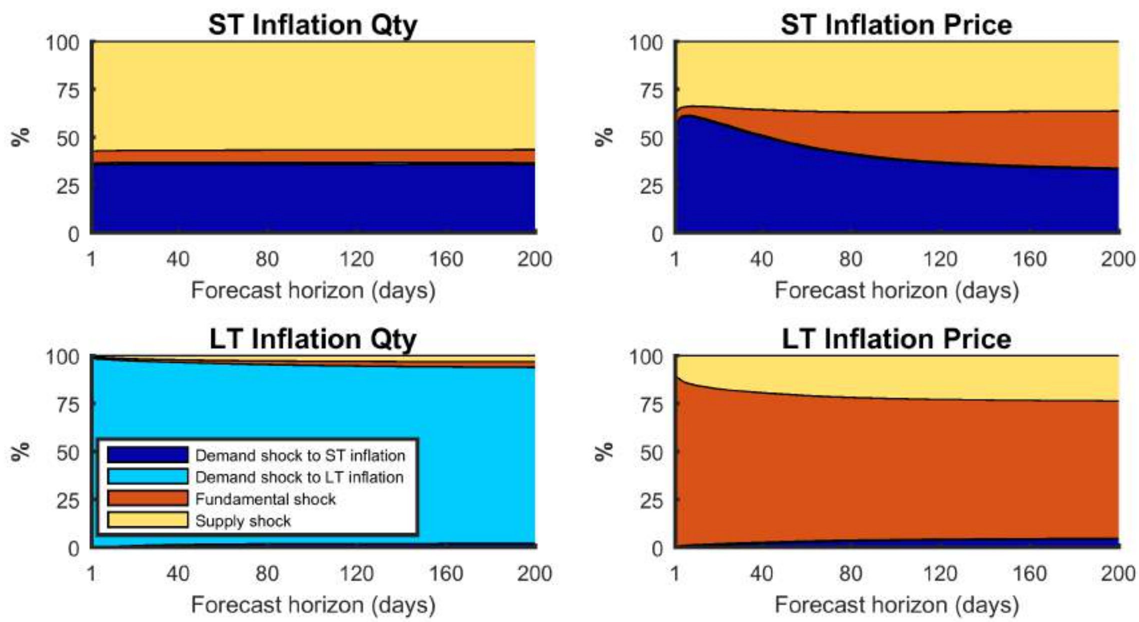


Note: Figure shows the median estimate from a total of 10,024 importance sampler draws that arise from our Bayesian VAR. The demand and supply functions that lie within shaded areas are the 68% confidence intervals.

5.3 Variance decompositions: the drivers of inflation swap prices

Figure 14 shows the forecast error variance decompositions for the different shocks. We plot these at different forecasts horizons but, confirming the previous finding of quick responses to shocks that have persistent effects, the decomposition is almost the same at all horizons.

Figure 14 FORECAST ERROR VARIANCE DECOMPOSITION



In the short-horizon, liquidity shocks drive nearly *all* the variation in inflation swap rates. Interestingly, while the focus in the literature on liquidity premia is is often on the market makers, we find that supply shocks from dealer banks only account for about one third of the variance of prices. The remaining two thirds come instead from liquidity shocks affecting the demand from pension funds. When it comes to quantities, this ordering reverses, with liquidity shocks to dealer banks accounting for about 60% of the variance, while liquidity shocks to pension funds account for the remaining 40%. Since regulatory constraints are probably not the main driver of hedge fund behavior, these results suggest putting more work into how background income risk (in the case of hedge funds, the in- and outflows) correlates with inflation, or what other capacity constraints hedge funds face in their ability to take inflation risk (perhaps linked to compensation and governance).

At long horizons, there is a stark contrast between quantities and prices. The former are almost entirely driven by liquidity frictions affecting pension funds. Changes in regulations, in constraints to taking risk, or in contributions and payouts into the pension funds come with large shifts in the demand curve in the long horizon market. Consistent with the flat supply curve from dealers that we found in the previous sub-section, this has a large effect on quantities, but very little effect on prices.

In such a market, prices can only move if the supply curve shifts, which could happen either because of shocks to fundamentals or liquidity shocks to dealers. The variance decomposition shows that shocks to fundamentals, that shift both supply and demand, account for roughly four fifths of the variance, with the remainder due to liquidity frictions affecting dealer banks that shift the supply curve.

Focusing on fundamentals, these variance decompositions suggest that changes in the 10-year inflation swap rate are a good indicator of fundamental expected inflation, but swap rates from the short horizon market are not. For macroeconomic purposes, most movements in the inflation swap rates up to three years can be dismissed.

6 Estimates: expected inflation and historical decompositions

This section presents and discusses our new measure of expected inflation from inflation swap prices, cleaned from liquidity shocks between 2019 and 2023. Historical decompositions reveal the contributions of the unobserved shocks to the observed market prices, informing the macroeconomic debates during this time.

6.1 Frictionless inflation driven by fundamental shocks

Figure 15 shows, in red, the realized measure of p, a conventional measure of long term expected inflation. The blue line instead presents the price that deducts all the shocks to liquidity frictions from the realized price. During this sample, when expected inflation fell, the traditional uncleaned measures fell by more; when it rose, the realized prices rose by more. Conventional measures therefore overstate the fluctuations in long horizon inflation expectations.

This applies to the period form October of 2022 to February of 2023, a crucial time when monetary policy had to assess whether persistently high inflation was unanchoring inflation expectations. Our estimates suggest that the risk of inflation being away from target was overstated because of heightened liquidity premium.

For each of the long and short horizon prices, we also estimated time series regressions of the following specification:

$$\Delta x_t^* = \zeta + \phi \Delta x_t + \upsilon_t, \;\; x \in \{p, P\} \tag{25}$$

Following the notation from Section 3, Δp_t^* refers to the first difference of our median estimate of the *frictionless* long horizon inflation swap rate (i.e., the swap breakeven rate in the absence of the estimated liquidity premia), and Δp_t is simply the first difference of the *actual* swap breakeven rate. Our coefficient of interest is ϕ, which captures a "back-of-an-envelope" correlation that measures the extent to which actual swap rates overreact to news about the fundamental. In other words, given our estimate of ϕ one could use actual swap prices to infer the extent to which its fundamental component has risen or fallen. The remaining terms are the constant ζ, and υ_t that could be a serially correlated disturbance. Our estimates of ϕ are 0.885 and 0.119 for the long horizon and short horizon swap prices, respectively. Both estimates are also highly significant with a t-statistic of 56.33 and 7.53 that both reject the null hypothesis of zero.[30] Taken at face value, this suggests every percentage point increase in the observed long horizon swap prices is associated with an average of 0.885 percentage point increase in the fundamental expected inflation. In contrast, every percentage point increase in the actual swap prices is only associated with a 0.119 percentage point increase in the fundamental on average. Both of these estimates are also consistent with the results from our forecast error variance decompositions, where they show fundamentals accounting for roughly 75%-80% of the variation in long horizon swap prices and less than 30% of the variation in short horizon swap prices.

6.2 The COVID-19 period

The initial COVID-19 pandemic period between February to September 2020 includes the first cases of the virus and the period following the announcement of the WHO declaring COVID-19 as a pandemic. The fall in demand owing to repeated lockdowns led to fears of a depression and deflation.

The top panel in Figure 16 shows the contribution of fundamentals and liquidity shocks to the inflation swap prices in the long-horizon market during this period, which are reported in the bottom panel.[31] The fall in swap prices was driven both by lower expected inflation and by liquidity shocks. Looking at swap prices alone would lead one to overstate the deflationary forces during the spring and summer of 2020.

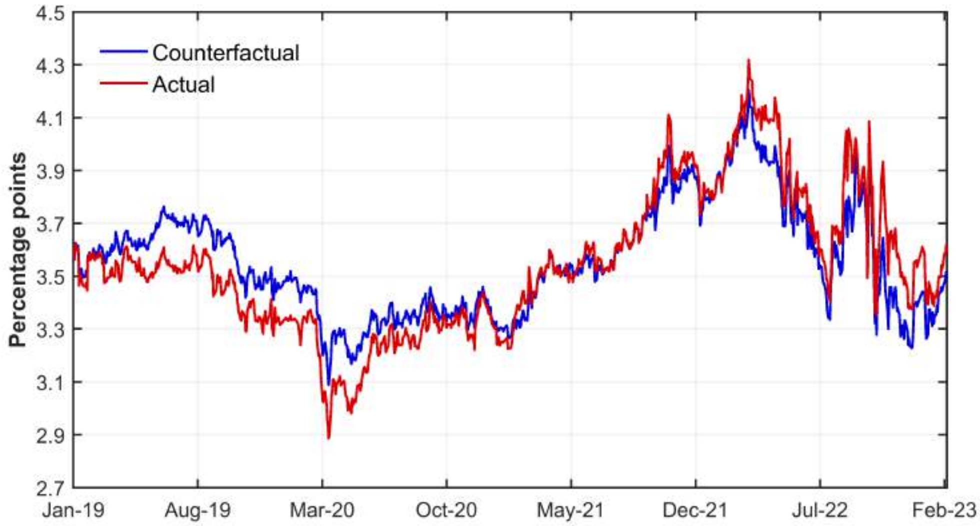
Looking at a few dates validates our decomposition. Fundamental expected inflation started to dive on 9^{th} March 2020 when the FTSE 100 fell by more than 8 percent, before reaching a trough on 11^{th} March 2020 when the WHO officially declared COVID-19 to be a global pandemic. This

[30] We use Newey-West standard errors with a maximum of 3 lags. Our estimates of ϕ are also robust to the addition of more lags of the independent variable. By including up to 5 lags of x_t, we obtain an estimate of ϕ to be 0.116 (with t-statistic of 7.44) and 0.879 (with t-statistic of 51.57) for the short horizon and long horizon swap prices respectively.

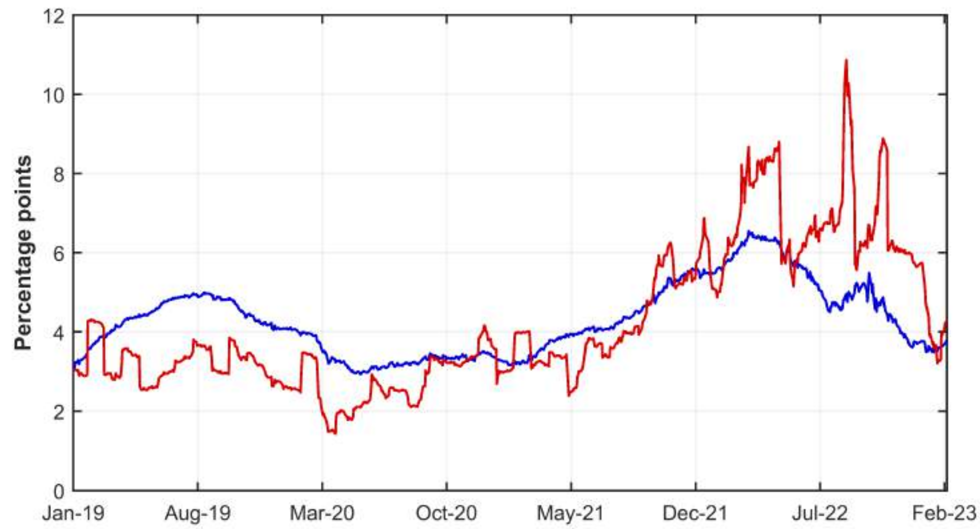
[31] The difference between the series in the bottom panel and the sum of the two series in the top panel is the contributions of the deterministic terms in the VAR.

Figure 15 FUNDAMENTAL EXPECTED INFLATION

(a) Long horizon UK RPI inflation swap rates



(b) Short horizon UK RPI inflation swap rates



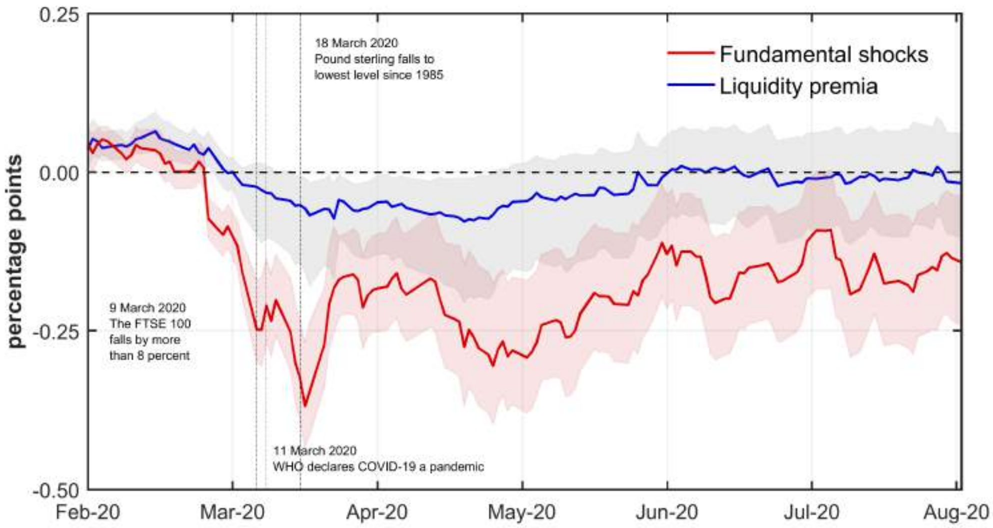
movement reversed on 18^{th} March 2020 when the Pound depreciated to its lowest exchange rate since 1985, providing an inflationary shock. Overall, the fact that 10-year fundamental expected inflation fell on average by only 20bps during this period shows a comfortable degree of anchoring.

6.3 The start of the Ukraine War

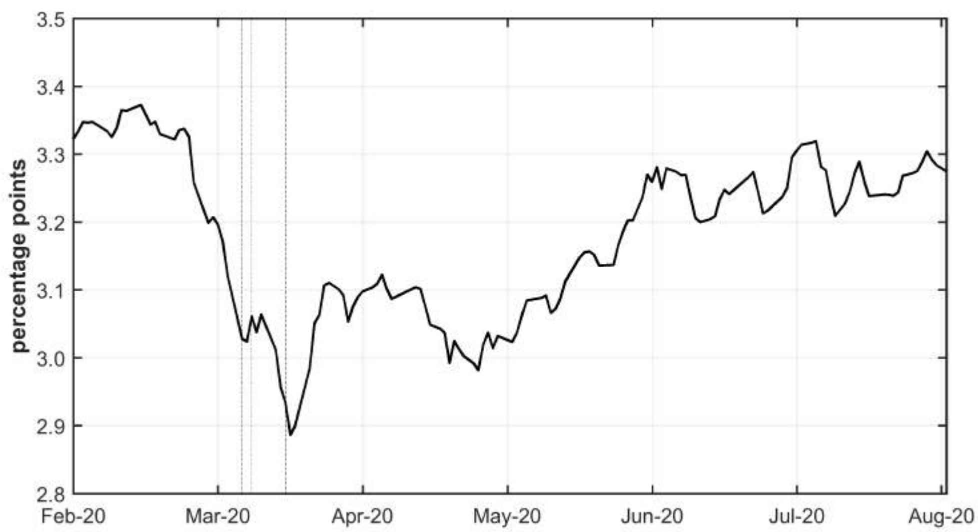
Another large shock to inflation during our sample was the start of the Ukraine War on the 24^{th} of February 2022. The months that followed came with large increases in the prices of crude oil and energy in the UK, and a re-evaluation of geopolitical constraints on trade. The price of inflation

Figure 16 THE COVID-19 PANDEMIC: SWAP RATES, LIQUIDITY, AND FUNDAMENTALS

(a) Estimated liquidity premia and fundamental shocks



(b) Long horizon UK RPI swap rates: COVID-19 period

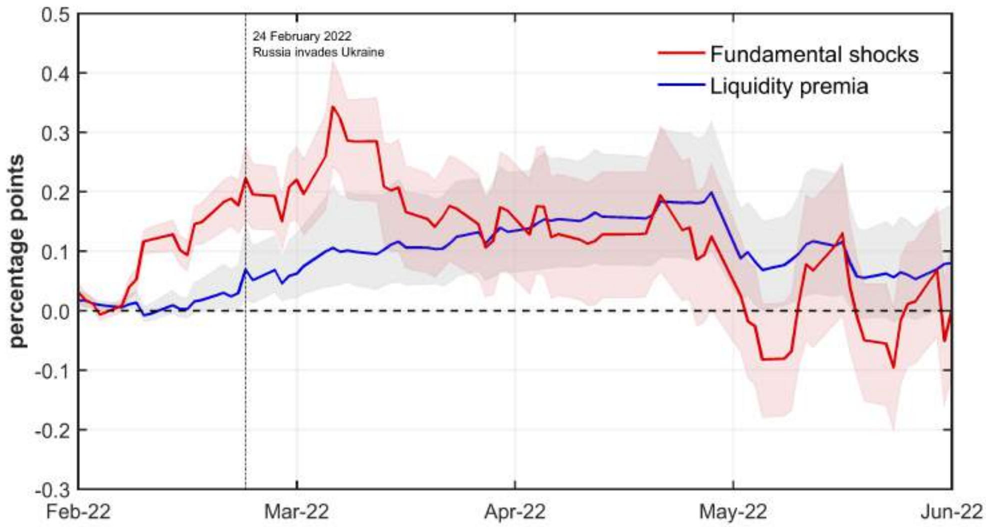


swaps rose sharply as we can see in Figure 17.

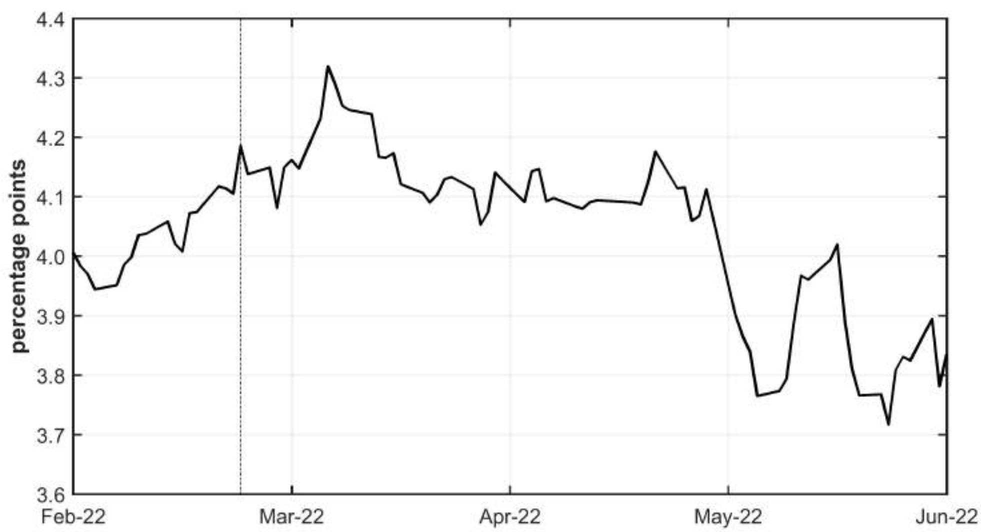
Again though, the rise in prices overstates the inflationary shock, because liquidity shocks were also elevating the prices by about 10bps during this period, as dealers constrained relative supply as demand grew. Our estimates are already able to pick up a rise in fundamental expected inflation by roughly 15bps in the lead-up to the war. Satellite information revealed heightened levels of Russian military activity at the border to Ukraine and financial markets were quick to incorporate the news. The date of the actual invasion provided a further increase of roughly the same magnitude.

Figure 17 The Ukraine War: swap rates, liquidity, and fundamentals

(a) Estimated liquidity premia and fundamental shocks



(b) Long horizon UK RPI swap rates: Ukraine War



6.4 The energy price guarantees and the UK LDI crisis

On the 6^{th} September 2022, Liz Truss became the UK's new Prime Minister having promised to tackle the UK's cost-of-living crisis brought about by the war in Ukraine. On the 6^{th} and 8^{th} September 2022, the government announced an "energy price guarantee", a price cap that substantially cut the effective prices that consumers would be paying for their household energy bills. This policy would have a large effect on measured headline RPI inflation in the following 12 to 24 months. On the 23^{rd} September 2022, the "Mini-Budget" was announced with large unfunded tax cuts. This fiscal expansion triggered a substantial fall in bond prices. This resulted in a sector-wide sell off of long-horizon bonds by liability-driven investment funds (the LDI in PFLDI) with further knock-on effects on wider financial stability. In order to stabilize the market, the Bank of England announced on 28^{th} September that it would temporarily buy a limited number of long-dated government bonds. One month later, on October 25^{th}, Rishi Sunak became Prime Minister and Jeremy Hunt the new Chancellor of the Exchequer. One of their first measures was to revert the tax cuts.

Figure 18 shows that in September and October of 2022, inflation swap prices were unusually volatile. The Truss energy cap reduced long-horizon inflation expectations, but the announcement of the fiscal expansion increased them sharply at first. This was quickly reversed as the Bank of England intervened, perhaps a reflection of successful communication with respect to the anchoring of expectations. One month later, when the tax cuts were reserved, expected inflation fell by approximately 20bps and stayed persistently lower. After the dust settled, this combination of shocks lowered long-horizon expected inflation.

Liquidity shocks pushed swap prices higher by more than 10bps, obscuring the fall in fundamentals. This contraction in supply by dealer banks may be a reflection of lasting concerns about the health of their counterparties in the PFLDI sector. Earlier, in September, the dominant liquidity shock came from the demand side, and pushed prices down. According to our estimates, PFLDIs were temporarily constrained in their ability to buy inflation swaps, which is consistent with the crisis in the sector during that month, and with the sector-wide deleveraging taking place in that month.

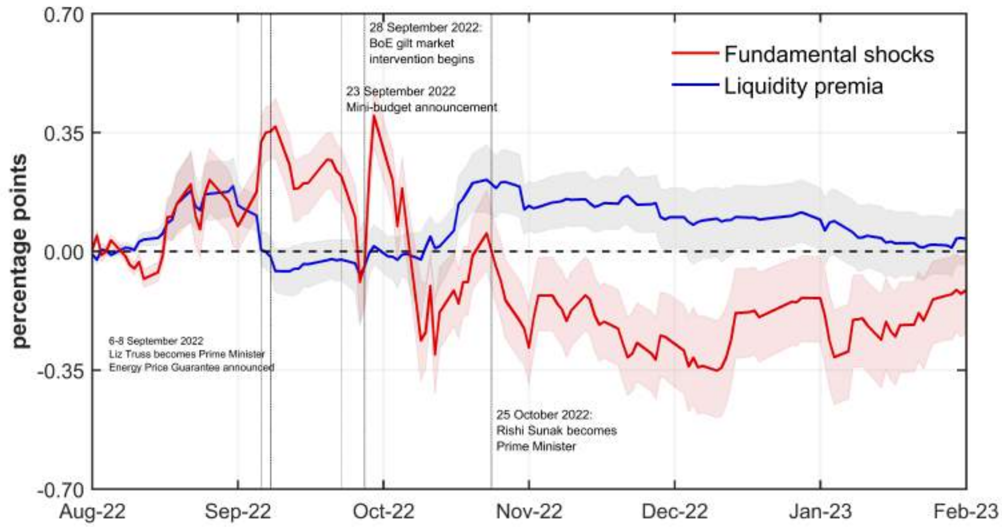
6.5 Squeezing signal out of short-term prices

So far, we have focussed on the long horizon market, since we found that this is the one that best captures fundamental shocks. While inflation swap prices in the short horizon market mostly reflects liquidity shocks, it can still provide some useful information. The top panel of Figure 19 shows the contribution of the estimated fundamental shocks to one-day ahead forecast errors while matching the local maxima in the swap prices to the most pertinent inflation news on that trading date. Even though liquidity shocks dominate, when large fundamental events happen, the inflation swap prices spike. So, at least for large events, they can carry some information about fundamental expectations of inflation.

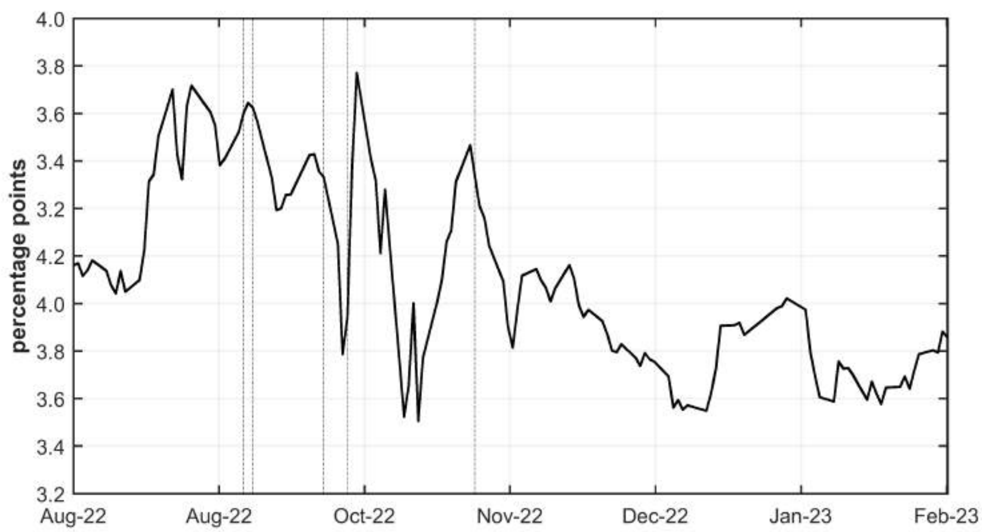
The bottom panel of Figure 19 focuses on the shocks to dealers' provision of inflation protection in the short horizon market. The variance decompositions suggested they account for 10-30% of the forecast error variance. There are two dates when they led to significant fluctuations in prices. The first is 6^{th} September 2022, when Liz Truss became British Prime Minister and was expected to unveil a package of energy price guarantees for households to cope with the rising energy cost.

Figure 18 AUTUMN 2022: SWAP RATES, LIQUIDITY, AND FUNDAMENTALS

(a) Estimated liquidity premia and fundamental shocks



(b) Long horizon UK RPI swap rates: UK's mini-budget

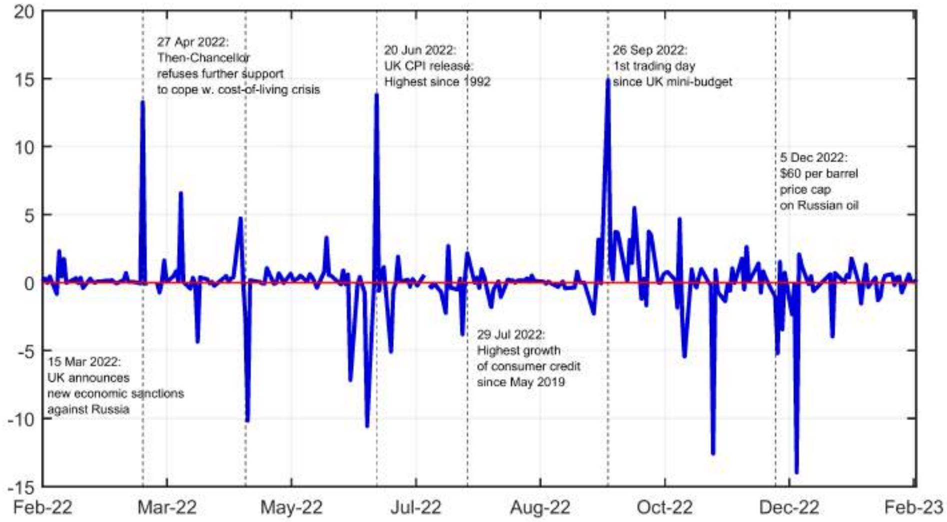


The second is 18^{th} October 2022, when the new government under Rishi Sunak reversed nearly all the tax cuts earlier proposed by Truss, a shock in the opposite direction.[32]

[32]Appendix E.13 shows the decompositions of both short horizon and long horizon prices into their fours shocks for the full sample.

Figure 19 Contribution of fundamental and liquidity shocks to 1-day ahead forecast error of short horizon inflation swap prices

(a) Contribution by fundamental shock (as ratio to all shocks)



(b) Contribution by dealers' supply shock

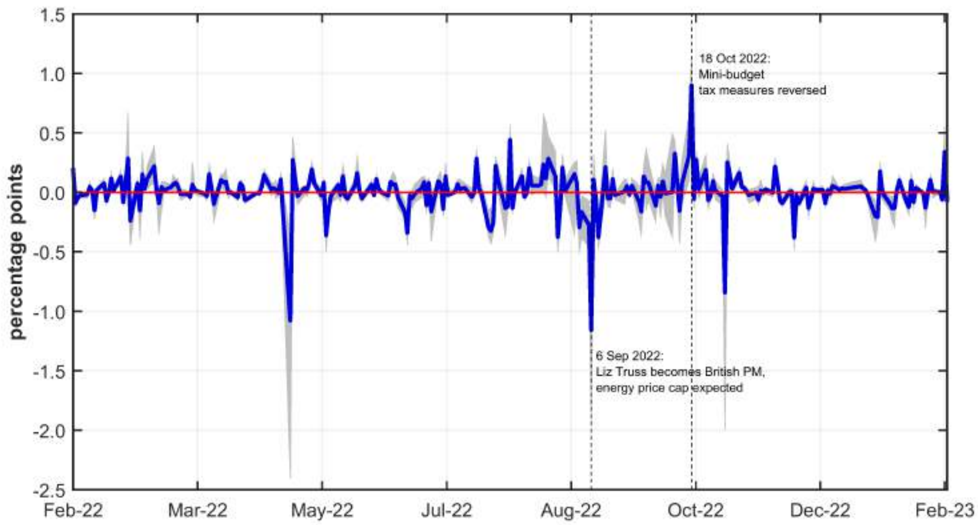
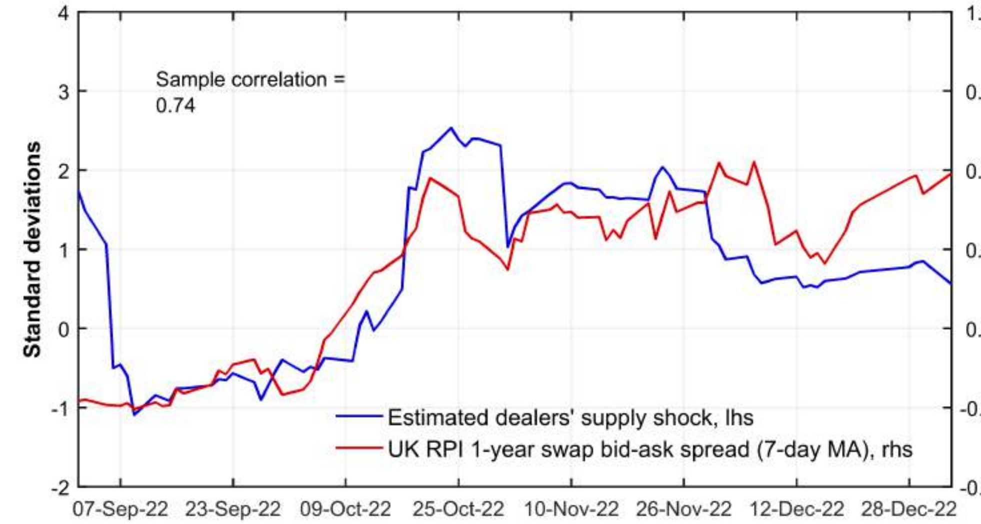
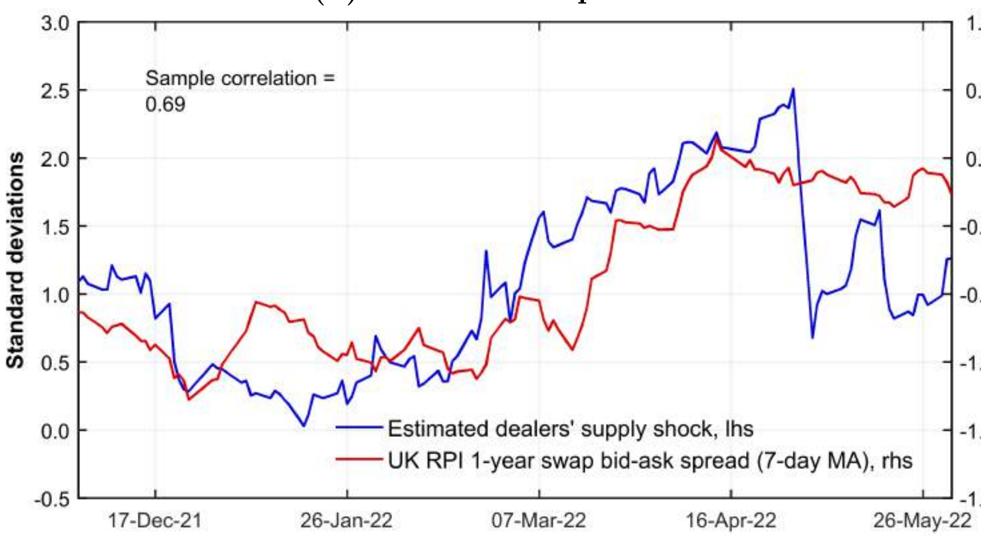


Figure 20 Comparisons with market bid-ask spreads

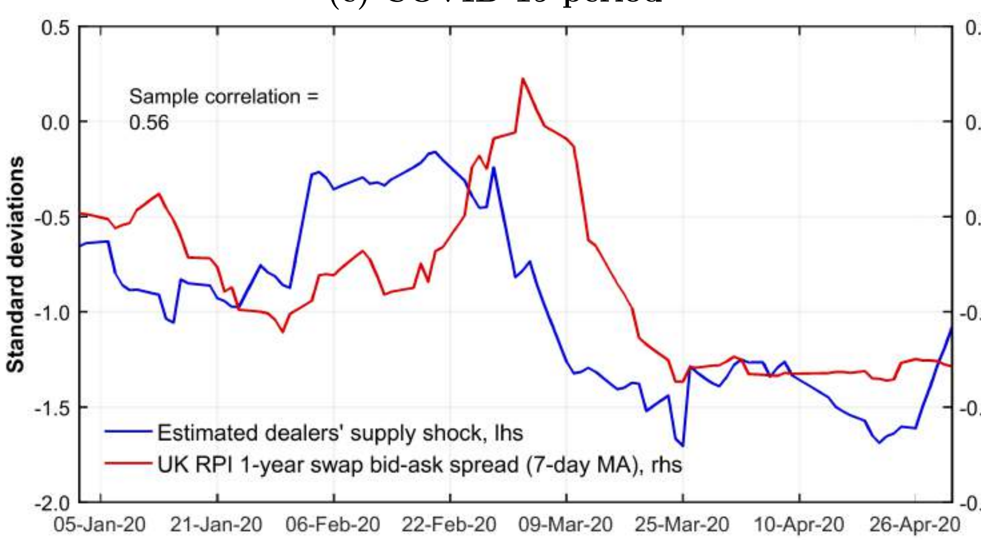
(a) LDI Crisis period



(b) Ukraine war period



(c) COVID-19 period



7 Estimates: whose beliefs drive expected inflation?

Fluctuations in the measures of expected inflation that have been purged from liquidity premia are driven by changes in beliefs about expected inflation and by changes in perceived risk premia. Our estimates do not allow us to separate subjective expected inflation form compensation for risk. This section proceeds by assuming that the fundamental shocks that we identified reflect shocks in expected inflation. Alternatively, we can interpret the estimates, still as capturing agents' beliefs and their dispersion and influence, but they are now beliefs about the sum of future inflation and future risk compensation. Under that assumption, we now measure the price variability that is driven by the beliefs of those demanding and those supplying inflation protection, and the dispersion of beliefs within these groups. Then, we contrast market measures of expected inflation with survey measures.

7.1 Individual beliefs of institutions

Using our estimate fundamental shocks driving frictionless expected inflation ($\varepsilon_{\pi,t}$), consider the following seemingly unrelated regressions:

$$\frac{q_{f,i,t}}{a_{f,i,t}} = constant + \beta_{f,i}\varepsilon_{\pi,t} + v_{f,i,t} \tag{26}$$

Recall the demand function in Proposition 1, and the solution for equilibrium prices in Lemma 1. They imply that the reduced-form coefficients $\beta_{f,i}$ would map into the structural coefficients of the model according to:

$$\beta_{f,i} = \frac{\mu_{f,i} - \Lambda}{\gamma_{f,i}\sigma_\pi^2(1-\rho_{\pi,d}^2)} \tag{27}$$

Therefore, across institutions, the coefficients the variation in the estimated $\beta_{f,i}$ will reflect the dispersion in their beliefs weighted by their risk aversion. If differences in risk aversion are small relative to differences in beliefs, then the cross-institution and cross-sectional variation in the estimates of $(\beta_{f,i}, \beta_{b,i}, \beta_{h,i})$ provides a measure of the dispersion in beliefs in these markets.

Intuitively, conditioning on the same fundamental shock, agents in this market will choose different positions to hold on inflation swaps if they disagree in their interpretation of what that shock implies for expected inflation. The more they disagree, the more differ the impact on their net position. If the dispersion in risk aversion goes in the opposite direction, in that those whose beliefs shift more are also the less risk averse, then this effect is amplified; if instead those whose beliefs shift more are more risk averse, then this effect is attenuated. Either way, the dispersion in responsiveness of trading strategies to the fundamental shock gives us a proxy estimate for how much disagreement there is in the market. This is an important object, both because disagreement drives trade, as well as because the literature on inflation expectations has found that its dispersion is an important diagnosis of whether expectations are drifting or anchored.[33]

Moreover, consider the impact of a counterfactual change in the inflation expectations of institution i on the price of the inflation swap, relative to the impact of a change in the inflation expectations of institution i'. From Lemma 1, this relative price impact is given by:

$$RelativePriceImpact_{i,i'} = \frac{\beta_{f,i} a_{f,i}}{\beta_{f,i'} a_{f,i'}}. \tag{28}$$

With the estimated $\beta_{f,i}$ and our measures of $a_{f,i}$, we can calculate this relative impact. It answers the question of which institutions are driving movements in expected inflation.

Figure 21 shows the estimates. The top panel shows the coefficients for dealers and hedge funds, with each estimate of $\beta_{b,i}$ and $\beta_{h,i}$, respectively, ordered from larger to smaller. Also in the figure, in a bold black line is the pooled coefficient estimate and the 95% confidence interval from clustering standard errors at the institutional level. We do not report the estimates for the pension funds because their dispersion is barely visible to the eye: whereas the cross-sectional standard deviation of coefficients is 0.052% for dealers and 0.017% for hedge funds, it is only 0.009% for pension funds. A first result is that there is a much larger dispersion of beliefs about inflation across dealers and hedge funds than there is across pension funds.

A second result, visible in the figure, is that while between dealers there are significant differences across all institutions, among hedge funds that trade inflation swaps, there are three that respond very strongly in one direction, and three that respond very strong in the other direction. This is consistent (and partly driven by) the large fluctuations in hedge fund positions, shifting from being net sellers to being net buyers of inflation protection.

At the same time, the estimated pooled coefficient for hedge funds is larger than for dealer funds. This maps into the differential reactiveness assumption that we made in one of the identification strategies. The same is true (and is reported in appendix G) in the long horizon market, where dealers witch sides having larger coefficients now relative to the pension funds.

Also in appendix G is a plot of the estimated coefficients against the size of the institution in the this market. There is little association with regards to the hedge funds, but for dealer banks, the larger ones are those that are more responsive to fundamentals. When inflation fundamentals change, it is the disagreement between these large dealer banks view and those in other sides of the market that leads to active trading and changes in equilibrium prices.

Panel (b) shows the price impacts in the short horizon market. When fundamentals change, the actions and beliefs of the majority of dealers and hedge funds has little impact on market prices. Three to five dealer banks, and three to five hedge funds respond a lot, in opposite directions, and

[33] See Mankiw, Reis, and Wolfers (2004) or Reis (2021).

this moves the prices of inflation swaps. Not only are the observed prices in short horizon inflation swaps mostly reflective of liquidity shocks, but even the part driven by fundamentals is driven by the views of ten or fewer institutions.

Panel (c) shows the price impact in the long horizon market. The dispersion among pension funds is significantly smaller. Correspondingly the demand contribution to the market price in long horizon market inflation expectations is a wider combination of the view of many different institutions. On the supply side, it is still the case that three banks have a larger impact than the others, but the difference of beliefs in relative terms is smaller than in the short horizon market.

7.2 Matching markets and surveys

In our data, we know the identities of the dealer banks. We can therefore map our estimated sensitivity of their beliefs to fundamentals $\beta_{b,i}$ to their name. In turn, these same dealer banks (with the exception of one of them) every month, during our sample, answered a survey ran by Bloomberg on their expected 1-year ahead RPI inflation. Regressing each institution's answer on the inflation swap price instrumenting with our fundamental shock, provides an estimate of their $\mu_{b,i}$. If the variability in risk aversion is small relative to the dispersion in beliefs, then these two separate estimates should line up.

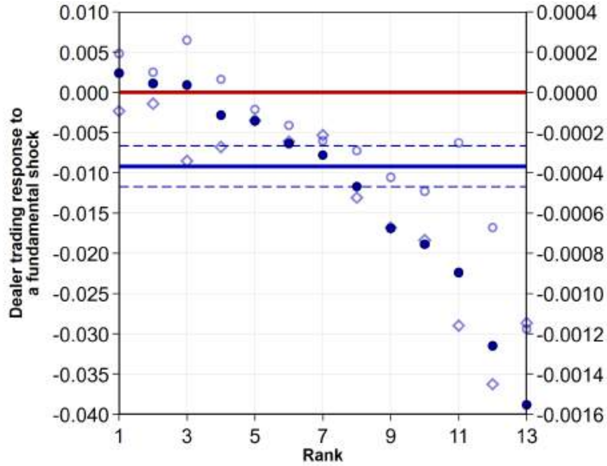
This provides an out-of-sample test of our estimates from the previous section, and an external validation of the conclusion we drew. The $\beta_{b,i}$ are backed out from the trading behavior of institutions in the inflation swap market. The $\beta_{b,i}$ instead map into their survey answers. Matching market behavior and survey answers at the micro level for inflation expectations has, to our knowledge, not been done before.

Figure 22 shows the scatter plots of these estimates. Each are in standardized units, the $\beta_{b,i}$ from market positions in the vertical axis, and the $\mu_{b,i}$ from survey answers in the horizontal axis. The association is clearly positive, even though the number of matched banks is small. This provides strong support for our estimates, as well as some support for using surveys of chief economists at these institutions as a signal of what drives the actual trading behavior of the institutions.

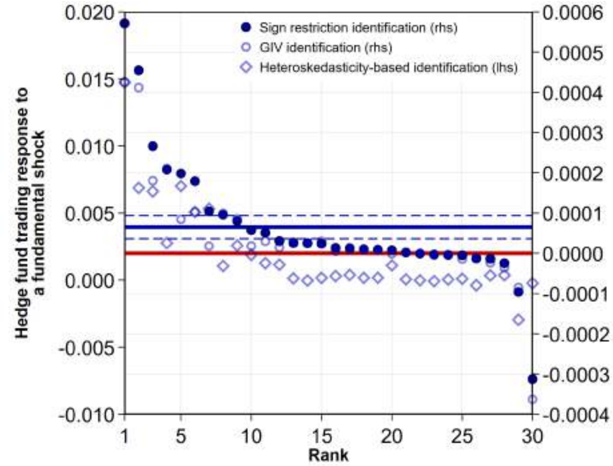
Figure 21 Estimates of the price impact of individual expectations

(i) Coefficients β_i ordered by magnitude

(a) Dealer banks (ST market)

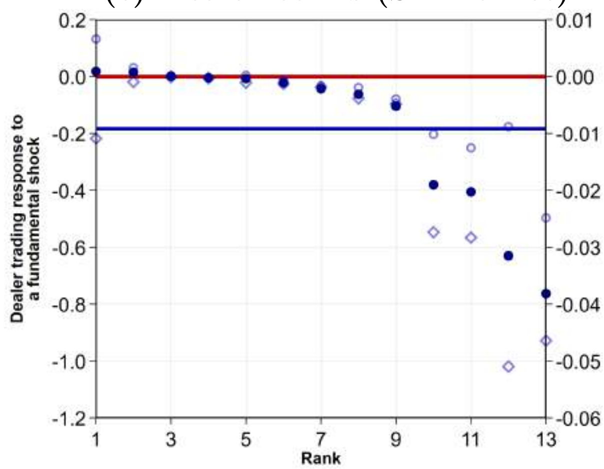


(b) Hedge funds

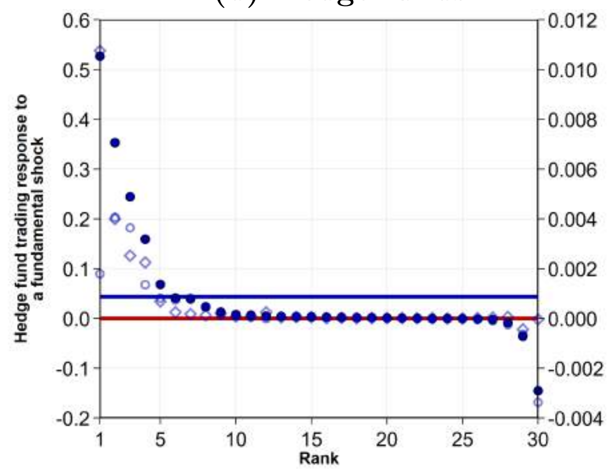


(ii) Price impact of institution $\beta_i a_i$ in the short horizon market

(c) Dealer banks (ST market)

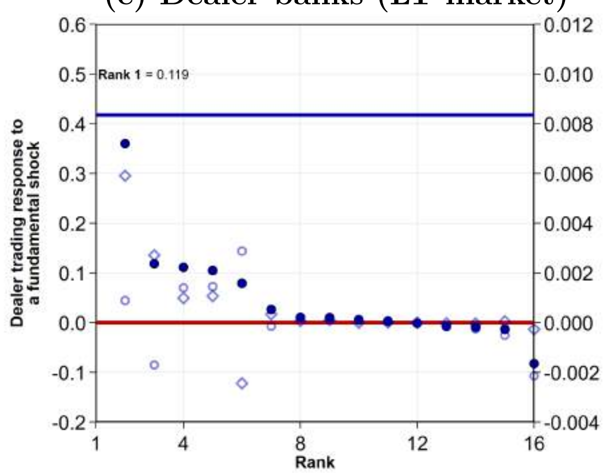


(d) Hedge funds

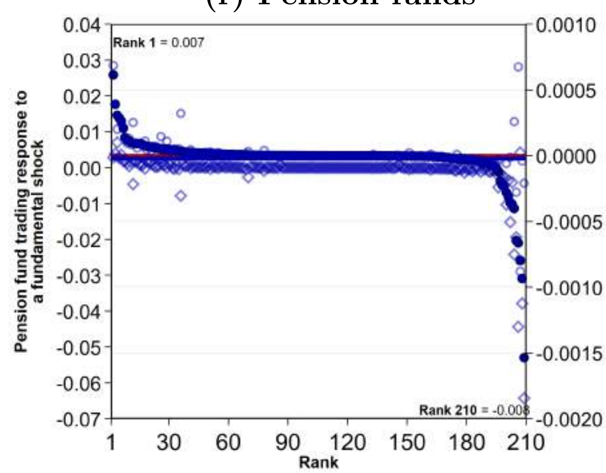


(iii) Price impact of institution $\beta_i a_i$ in the long horizon market

(e) Dealer banks (LT market)

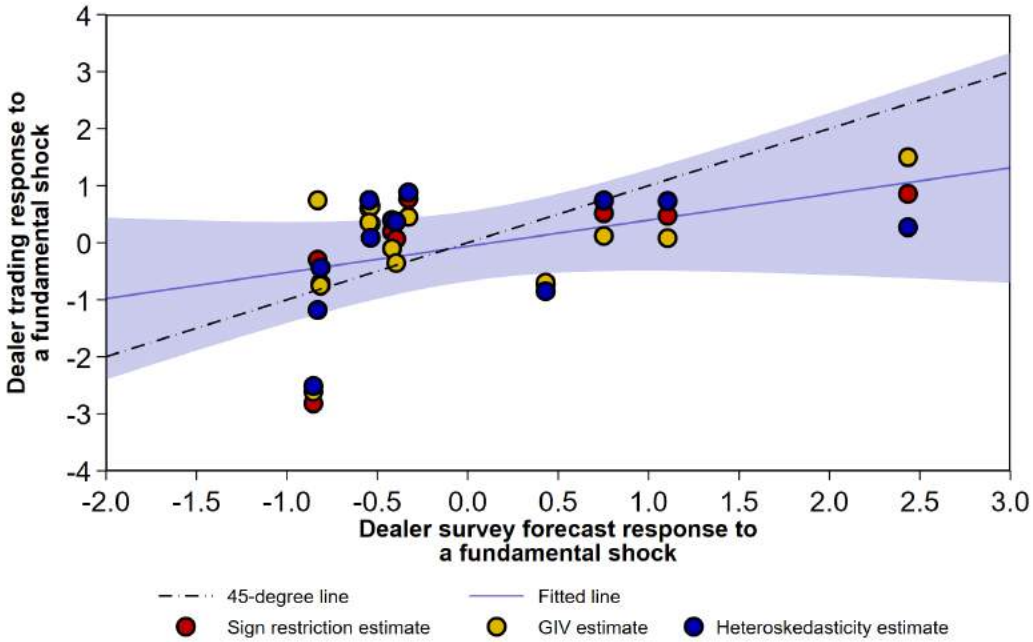


(f) Pension funds



Note: The thick blue lines are pooled coefficient estimates with standard errors clustered at the institutional-level (where each is identified by a legal entity identifier), dashed blue lines are their 95% confidence intervals. Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

Figure 22 Expectations and heterogeneity in trading



8 Conclusion

This paper complemented the data on prices of inflation swap contracts that are heavily used by researchers and policymakers with valuable data on the quantities behind the prices and the institutions engaging in trades. It then provided a model to decompose fluctuations in this market into liquidity premia and fundamentals, and further separate fundamentals into the dispersion of beliefs behind them. Finally, it provided new identification strategies to identify the demand and supply functions in this large market in which inflation risk is traded, and the shocks that drive them. This analysis led to the following lessons about inflation risk:

First, there is a remarkable segmentation of the market across horizons. In short horizons, hedge funds and dealers are active, alternating between negative and positive net positions that average to zero. In long horizons, dealers steadily provide inflation protection to pension funds. Therefore, dealers provide the link between the two markets by supplying swaps to both.

Second, because we have high-frequency trading data at the institution-level, for four years with many news, we can identify the four supply and demand functions in this market. Three separate identification strategies exploit different dimensions of the data to do so. Reassuringly, the three yield similar empirical estimates, suggesting that the results are robust.

Third, in the market, prices seem to fully reflect information after one to three days. There is some persistence, so markets are not fully efficient, but they are quite close to it, perhaps because investors in this market are very sophisticated, although subject to different constraints and operating with different beliefs.

Fourth, the slope of the supply function of dealer banks is close to horizontal at long horizons (but not so at short horizons). Therefore, fluctuations in quantities traded in this market reflect almost entirely liquidity shocks shifting the demand from pension funds. Whether this is due to institutional characteristics, or due to the relative bargaining power of different clients, is unclear but future research can explore this intriguing result.

Fifth, fundamental shocks drive the long-horizon swap prices, while liquidity shocks drive the short-horizon prices. Therefore, even though one-year inflation swap prices are sometimes used (amongst other measures) to quantify expected inflation and to guide monetary policy, our estimates suggest that much of the variation in these prices can be explained by liquidity shocks to hedge funds and dealers. Comparing our estimates with key dates of market dysfunction confirms this view.

Sixth, we produce a measure of expected inflation at longer horizons that is cleaned of liquidity frictions. This measure shows that expected inflation is more stable than prices alone suggest. At the start of 2023, this measure paints a more optimistic picture of the anchoring of inflation expectations relative to conventional measures. Looking at three recent episodes that ignited debates about inflation and monetary policy, our measure shows that the fears of deflation were overstated during the pandemic, the fears of runaway inflation were overstated during the energy crisis, and that the fiscal measures of Autumn 2022 had as much of an effect on swap prices through liquidity premia as they did through expected inflation.

Seventh, we found that there is significant dispersion in the beliefs about inflation within and between dealer banks and hedge funds. In the short horizon market, a handful of institutions have a large price impact and dominate the formation of prices. In the long horizon market, pension funds have closer beliefs and their price impact is more homogeneous. The beliefs about inflation of dealer banks inferred from trading activity and from surveys line up remarkably well.

Overall, this paper contributed a series of data for measurement, techniques for estimation, and empirical results that leave some intriguing open challenges to the finance literature on the mechanics and segmentation of an important financial market, the macroeconomic literature on fluctuations in expected inflation, and the behavioral literature on dispersion in beliefs that, we hope, will inspire future research.

References

Abad, J., I. Aldasoro, C. Aymanns, M. D'Errico, L. F. Rousová, P. Hoffmann, S. Langfield, M. Neychev, and T. Roukny (2016): " Shedding light on dark markets: First insights from the new EU-wide OTC derivatives dataset," *ESRB: Occasional Paper Series No. 11*.

Andrade, P. and F. Ferroni (2021): "Delphic and odyssean monetary policy shocks: evidence from the euroa area," *Journal of Monetary Economics*, 117, 816–832.

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.

Bai, J. (2009): "Panel Data Models With Interactive Fixed Effects," *Econometrica*, 77, 1229–1279.

Bai, J. and S. Ng (2002): "Determining the number of factors in approximate factor models." *Econometrica*, 70(1), 191–221.

Beechey, M. J., B. K. Johannsen, and A. T. Levin (2011): "Are Long-Run Inflation Expectations Anchored More Firmly in the Euro Area Than in the United States?" *American Economic Journal: Macroeconomics*, 3, 104–29.

Begenau, J., M. Piazzesi, and M. Schneider (2015): "Banks' risk exposures," *NBER Working Paper No. 21334*.

Benos, E., G. Ferrara, and A. Ranaldo (2022): "Collateral Cycles," *Bank of England Staff Working Paper No. 966*.

Boons, M., F. Duarte, F. de Roon, and M. Szymanowska (2020): "Time-varying inflation risk and stock returns," *Journal of Financial Economics*, 136, 444–470.

Brunnermeier, M., D. Palia, K. A. Sastry, and C. A. Sims (2021): "Feedbacks: Financial Markets and Economic Activity," *American Economic Review*, 111, 1845–79.

Campbell, J. Y., C. Pflueger, and L. M. Viceira (2020): "Macroeconomic Drivers of Bond and Equity Risks," *Journal of Political Economy*, 128, 3148–3185.

Cenedese, G., P. Della Corte, and T. Wang (2021): "Currency Mispricing and Dealer Balance Sheets," *The Journal of Finance*, 76, 2763–2803.

Cenedese, G., A. Ranaldo, and M. Vasios (2020): "OTC premia," *Journal of Financial Economics*, 136, 86–105.

Cieslak, A. and C. Pflueger (2023): "Inflation and Asset Returns," CEPR discussion paper 18003.

Covi, G., J. Brookes, and C. Raja (2022): "Measuring Capital at Risk in the UK banking sector: a microstructural network approach," *Bank of England Staff Working Paper No. 983*.

Czech, R., P. Della Corte, S. Huang, and T. Wang (2022): "FX Option Volume," *Bank of England Staff Working Paper No. 964*.

D'Amico, S. and T. B. King (2023): "Inflation hedging products," in *The Research Handbook of Financial Markets*, ed. by R. Gurkaynak and J. Wright, Edward-Elgar.

Fang, X., Y. Liu, and N. Roussanov (2022): "Getting to the Core: Inflation Risks Within and Across Asset Classes," Working Paper 30169, National Bureau of Economic Research.

Ferrara, G., P. Mueller, G. Viswanath-Natraj, and J. Wang (2022): "Central bank swap lines: micro-level evidence," *Bank of England Staff Working Paper No. 977.*

Gabaix, X. (2016): "Power laws in economics: An introduction," *Journal of Economic Perspectives*, 30, 185–206.

Gabaix, X. and R. S. Koijen (2020): "Granular instrumental variables," *National Bureau of Economic Research Working Paper Series.*

——— (2021): "In search of the origins of financial fluctuations: The inelastic markets hypothesis," *National Bureau of Economic Research Working Paper Series.*

Gabaix, X., R. S. Koijen, F. Mainardi, S. S. Oh, and M. Yogo (2022): "Asset Demand of U.S. Households," Harvard Umniversity manuscript.

Hanson, S., A. Malkhozov, and G. Venter (2022): "Demand and Supply Imbalance Risk and Long-Term Swap Spreads," *Unpublished working paper.*

Hau, H., P. Hoffmann, S. Langfield, and Y. Timmer (2021): "Discriminatory Pricing of Over-the-Counter Derivatives," *Management Science*, 67, 6660–6677.

Haubrich, J., G. Pennacchi, and P. Ritchken (2012): "Inflation Expectations, Real Rates, and Risk Premia: Evidence from Inflation Swaps," *The Review of Financial Studies*, 25, 1588–1629.

Jiang, E., G. Matvos, T. Piskorski, and A. Seru (2023): "Limited Hedging and Gambling for Resurrection by U.S. Banks During the 2022 Monetary Tightening?" *Unpublished Working Paper.*

Khetan, U., I. Neamtu, and I. Sen (2023): "The market for sharing interest rate risk: quantities behind prices," Bank of England Staff Working Paper 1,031, Bank of England.

Koijen, R. S., F. Koulischer, B. Nguyen, and M. Yogo (2021): "Inspecting the mechanism of quantitative easing in the euro area," *Journal of Financial Economics*, 140, 1–20.

Koijen, R. S., R. J. Richmond, and M. Yogo (2020): "Which investors matter for equity valuations and expected returns?" *National Bureau of Economic Research Working Paper Series.*

Koijen, R. S. and M. Yogo (2020): "Exchange rates and asset prices in a global demand system," *National Bureau of Economic Research Working Paper Series.*

Koijen, R. S. J. and M. Yogo (2019): "A Demand System Approach to Asset Pricing," *Journal of Political Economy*, 127, 1475–1515.

Lütkepohl, H., M. Meitz, A. Netšunajev, and P. Saikkonen (2021): " Testing identification via heteroskedasticity in structural vector autoregressive models," *The Econometrics Journal*, 24, 1–2.

Mankiw, N. G., R. Reis, and J. Wolfers (2004): *Disagreement about Inflation Expectations*, NBER Macroeconomics Annual 2003, vol. 18, 209–248.

Mann, C. L. (2022): "Inflation expectations, inflation persistence, and monetary policy strategy," *Speech at the 53rd Annual Conference of the Money Macro and Finance Society, University of Kent.*

McPhail, L., P. Schnabl, and B. Tuckman (2023): "Do Banks Hedge Using Interest Rate Swaps?" *NBER Working Paper No. 31166*.

Ramsden, D. (2022): "Shocks, inflation, and the policy response," *Speech at the at Securities Industry Conference*.

Reis, R. (2020): "The People versus the Markets: A Parsimonious Model of Inflation Expectations," CEPR discussion paper 15624.

——— (2021): "Losing the Inflation Anchor," *Brookings Papers on Economic Activity*, 307–361.

Stock, J. H. and M. W. Watson (2018): "Identification and Estimation of Dynamic Causal Effects in Macroeconomics Using External Instruments," *The Economic Journal*, 128, 917–948.

Vayanos, D. and J.-L. Vila (2021): "A Preferred-Habitat Model of the Term Structure of Interest Rates," *Econometrica*, 89, 77–112.

Appendix

A Data: additional information

We provide additional information on the data used in the empirical analysis in this section of the Appendix.

A.1 Trade repository data and the data cleaning procedures

We begin with 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.

The raw data files: 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 31^{st} 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.[34]

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

[34] `https://www.fca.org.uk/markets/uk-emir/reporting-obligation`

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.[35] 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.[36] 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 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.

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.

Data cleaning procedure. 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

[35] 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.

[36] 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.

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".[37] 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.[38] 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.[39] 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.[40]

[37]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").

[38]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.

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

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

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 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.[41] 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.[42] 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.[43]

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

[41] 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.

[42] 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.

[43] 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 31^{st} December 2020.[44] 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.[45] Subsequently, we drop all swap transactions that were not confirmed according to Article 12 of Commission Delegated Regulation No. 149/2013.[46] We take these steps to be a conservative approach 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"

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

[45] https://www.fca.org.uk/markets/uk-emir/reporting-obligation

[46] https://www.legislation.gov.uk/eur/2013/149/article/12

should be categorized as UK CPI inflation swaps.[47] 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.[48]

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.[49] 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.

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 above-mentioned 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.[50] 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

[47] 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".

[48] 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.

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

[50] 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.

use to construct a monthly time series of both gross and net notional positions.[51] 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.[52] 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 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.[53] 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.[54] 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

[51]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.

[52]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.

[53]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.

[54]These swap prices have been adjusted for RPI indexation lags. See Appendix A.2 for a formal description of the problem.

maturity dates and effective dates.[55] 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).[56] 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 with the 30-year swap rate.

9. **Constructing daily price-quantity pairs in the long and short horizon markets**. The dataset is further restricted to dealer-client trades 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 horizon 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 horizon 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 its 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 horizon 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.

Finally, noting that data quality is poor prior to 2019, we further dropped all the corresponding quantities and prices with an execution date that precedes 1st January 2019. This yields a time series dataset of 879 x 4 observations (ST quantity, ST price, LT quantity, LT price) that is the raw data matrix used for VAR estimation in implementing all of our three identification strategies.

A.2 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.

We begin by noting that there are two relevant market conventions for RPI swap pricing. First, the floating rate index on which liabilities of the floating payer are calculated is referenced to the RPI index from 2-months ago. This is known as the RPI indexation lag. Take a 1-year zero coupon RPI swap for consideration. 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. However, because the May (and

[55] 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.

[56] 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.

possibly June release) 2023 RPI would have been released at the trade date, the inflation swap seller would require a breakeven rate to at least compensate for her liability arising from the growth in the RPI index between April 2023 to May 2023 (or June 2023). This is the known component of the breakeven price. The forward-looking component of this price would thus be attributed to the expected change in the RPI index between April 2024 and June 2023, thus reflecting market's expectation of RPI inflation 10-months from now. 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.

Formally, let $p^{(N)}_{t_d,t_m}$ denote the annualised breakeven rate of a N-year zero coupon RPI swap that is traded in date t_d of month t_m. As explained above, the two time indexes are required because inflation swaps trade at a daily frequency, thus swap prices vary from day-to-day reflecting changes in the market's expectation of the future RPI given possibly new information of the economy. At the same time, the reference RPI index that is tied to the floating leg changes only at a monthly frequency.

Using the definition of a swap breakeven rate, this implies that at the trade date of t_d in month t_m it must satisfy:

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

where $RPI^e_{\cdot,t_m+12N-2}$ denotes the date t_d expectation (in month t_m) of the future RPI realisation in month $t_m+12N-2$, conditional on the information of the economy at date t_d. That is, $RPI^e_{\cdot,t_m+12N-2} = \mathbb{E}[RPI_{\cdot,t_m+12N-2}|\mathcal{I}_{t_d}]$. The equation above can be rewritten as:

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

Using this equation, we are able to 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, however, poses a further set of complications that is worth mentioning. In particular, 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. Our indexation-adjustment is based on the premise 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 annualised rate reflecting RPI inflation expectations over the horizon from month t_m (inclusive) to month $t_m+12N-2$. We subsequently use the following equation to implement the indexation-adjustment:

$$(1+p^{(N)}_{t_d,t_m})^N = \underbrace{\left(\frac{RPI^e_{\cdot,t_m+12N-2}}{RPI_{\cdot,t_m-1}}\right)}_{unknown\ payoff}\underbrace{\left(\frac{RPI_{\cdot,t_m-1}}{RPI_{\cdot,t_m-2}}\right)}_{known\ payoff} \tag{A.3}$$

We let $x^{(N)}_{t_d,t_m}$ denote the indexation-adjusted swap rate of the N-year zero coupon swap. Expressing it in an annualised term, we have:

$$(1+x^{(N)}_{t_d,t_m})^N = \left(\frac{RPI^e_{\cdot,t_m+12N-2}}{RPI_{\cdot,t_m-1}}\right)^{\frac{12N}{12N-1}} \tag{A.4}$$

Substituting Equation (A.4) into Equation (A.3) yields:

$$(1 + p_{t_d,t_m}^{(N)})^N = \left[(1 + x_{t_d,t_m}^{(N)})^N\right]^{\frac{12N-1}{12N}} \left(\frac{RPI_{\cdot,t_m-1}}{RPI_{\cdot,t_m-2}}\right) \tag{A.5}$$

Finally, taking logs and rearranging, we solve for the indexation-adjusted N-year zero coupon swap breakeven rate $x_{t_d,t_m}^{(N)}$, for $N = 1, ..., 10, 12, 15, 20, 25, 30, 35, 40, 45, 50$:

$$\log(1 + x_{t_d,t_m}^{(N)}) = \frac{12N}{12N-1}\Big[\log(1 + p_{t_d,t_m}^{(N)}) - \frac{1}{N}\log\big(\frac{RPI_{\cdot,t_m-1}}{RPI_{\cdot,t_m-2}}\big)\Big] \tag{A.6}$$

A.3 Variables

Our 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 horizon of the contracts per year. We aggregate contract horizons into three buckets: three years or less Q, 10 years or more q, and the remainder.

The two key observables used in estimation then are a balanced panel for $q_{f,i,t}$ and $a_{f,i,t}$ (and same for b and h). The trading activity is 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 PFLDI 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 horizon 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 PFLDI institutions 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.

A.4 Additional summary statistics

In this section, we provide additional summary statistics for a single trade state report, covering all outstanding trades in the UK RPI inflation swap market on the last day in our sample: 10^{th} February 2023. As shown in Table A.1, we observe the largest number of trades by pension funds, followed by insurers, LDI funds and hedge funds. Moreover, we find the largest average trades size 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 A.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									
PFLDIs	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: All columns except the last are in units of USD millions. 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: 10 February 2023 DTCC OTC interest rate trade state file.

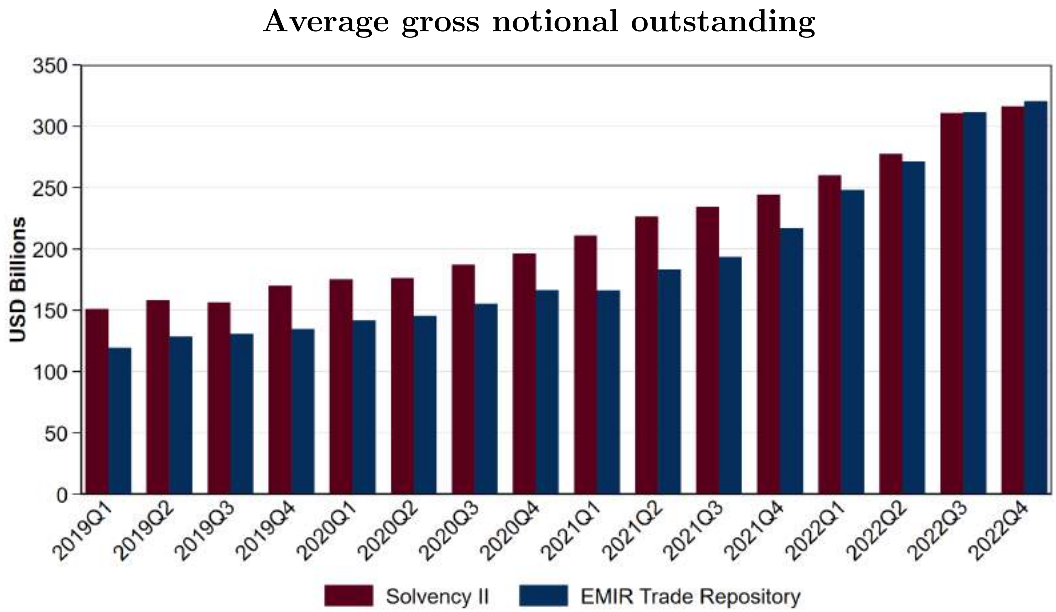
B A robustness check: comparison to alternative data

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 B.1 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.[57] Throughout our sample period, the EMIR TR data cover almost 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.

[57] Section F.1 of the Appendix 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.

Figure B.1 Comparison of Solvency II insurance holdings and EMIR TR data



Source: Solvency II data and DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

C Proofs of the model results

This section provides the proofs of the two propositions and the two lemmas.

C.1 Proof of proposition 1

Facing uncertainty on expected inflation, market returns and background risk, an individual pension fund believes they are jointly normally distributed with mean $\boldsymbol{\nu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$:

$$\boldsymbol{\nu} = \begin{pmatrix} \mathbb{E}_{f,i}[\pi] - p \\ \mathbb{E}_{f,i}[d] - s \\ \mathbb{E}_{f,i}[y_{f,i}] \end{pmatrix} = \begin{pmatrix} \mu_{f,i}\pi^e - p \\ \theta_d - s \\ 0 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\pi^2 & \sigma_{\pi,d} & \sigma_{\pi,y_{f,i}} \\ \sigma_{d,\pi} & \sigma_d^2 & 0 \\ \sigma_{y_{f,i},\pi} & 0 & \sigma_{y_{f,i}}^2 \end{pmatrix} \tag{C.1}$$

Let $\Delta a_{f,i} = a'_{f,i} - a_{f,i}$, the change in the market value of an institution's portfolio overnight. Since $\Delta a_{f,i} = (\pi - p)q_{f,i} + (d - s)e_{f,i} + y_{f,i}$ is a linear combination of Gaussian random variables it is also Gaussian with mean and variance:

$$\mathbb{E}_{f,i}[\Delta a_{f,i}] = (\mu_{f,i}\pi^e - p)q_{f,i} + (\theta_d - s)e_{f,i} \tag{C.2}$$

$$\begin{aligned} Var_{f,i}[\Delta a_{f,i}] = {} & q_{f,i}^2 Var[\pi - p] + e_{f,i}^2 Var[d - s] + Var[y_{f,i}] \\ & + 2q_{f,i}e_{f,i}Cov[(\pi - p)(d - s)] + 2q_{f,i}Cov[(\pi - p)y_{f,i}] \end{aligned} \tag{C.3}$$

Replacing this budget constraint into the the pension fund's objective function and using normality:

$$\mathbb{E}_{f,i}\Big[- exp\big(- \widetilde{\gamma}_{f,i}a'_{f,i}\big)\Big] = -exp\left\{ -\widetilde{\gamma}_{f,i}\left[a_{f,i} + \mathbb{E}_{f,i}[\Delta a_{f,i}] - \frac{\widetilde{\gamma}_{f,i}Var_{f,i}[\Delta a_{f,i}]}{2} \right] \right\} \tag{C.4}$$

The pension fund's optimization is then equivalent to a mean-variance problem

$$\max_{\{q_{f,i},e_{f,i}\}} (\mu_{f,i}\pi^e - p)q_{f,i} + (\theta_d - s)e_{f,i} - \frac{\widetilde{\gamma}_{f,i}}{2}\Big[q_{f,i}^2\sigma_\pi^2 + e_{f,i}^2\sigma_d^2 + \sigma_{y_{f,i}}^2 + 2q_{f,i}e_{f,i}\sigma_{\pi,d} + 2q_{f,i}\sigma_{\pi,y_{f,i}}\Big]$$
$$s.t.\;\; G_f(q_{f,i}, z_{f,i}) \geq 0 \tag{C.5}$$

Taking market prices as given, the first-order conditions necessary to attain a maximum are given by:

$$q_{f,i}^* = \frac{\mu_{f,i}\pi^e - p^*}{\widetilde{\gamma}_{f,i}\sigma_\pi^2} - \frac{\sigma_{\pi,d}}{\sigma_\pi^2}e_{f,i}^* - \frac{\sigma_{\pi,y_{f,i}}}{\sigma_\pi^2} - \frac{\lambda_{f,i}G_f'(q_{f,i}^*, z_{f,i})}{\widetilde{\gamma}_{f,i}\sigma_\pi^2} \tag{C.6}$$
$$e_{f,i}^* = \frac{\theta_d - s^*}{\widetilde{\gamma}_{f,i}\sigma_d^2} - \frac{\sigma_{\pi,d}}{\sigma_d^2}q_{f,i}^*, \tag{C.7}$$

where $\lambda_{f,i}$ is the Lagrange multiplier associated with institution i's capacity constraint that satisfies the two-part Kuhn-Tucker conditions:

$$\lambda_{f,i}G_f(q_{f,i}^*, z_{f,i}) = 0 \quad \text{and} \quad G_f(q_{f,i}^*, z_{f,i}) \geq 0 \quad \text{and} \quad \lambda_{f,i} \geq 0, \tag{C.8}$$

and $q_{f,i}^*$ denotes the institution's optimal portfolio allocation of inflation swap contracts in equilibrium.

Combining the First-Order conditions above, and substituting the definition $\widetilde{\gamma}_{f,i} = \gamma_{f,i}/a_{f,i}$, the solution for $q_{f,i}^*$ as a function of the model's primitives is

$$\frac{q_{f,i}^*}{a_{f,i}} = \frac{\mu_{f,i}\pi^e - p^*}{\gamma_{f,i}\sigma_\pi^2(1-\rho_{\pi,d}^2)} - \frac{\sigma_d}{\sigma_\pi}\left[\frac{\theta_d - s^*}{\gamma_{f,i}\sigma_d^2(1-\rho_{\pi,d}^2)}\right]\rho_{\pi,d} - \left[\frac{1}{(1-\rho_{\pi,d}^2)\sigma_\pi^2}\right]\left[\frac{\sigma_{\pi,y_{f,i}}}{a_{f,i}} + \frac{\lambda_{f,i}g_f(q_{f,i}^*, z_{f,i})}{\gamma_{f,i}}\right] \tag{C.9}$$

where $\rho_{\pi,d}^2 = \sigma_{\pi,d}/\sigma_\pi\sigma_d$ is the correlation between π and d. This completes the proof of proposition 1.

C.2 Proof of proposition 2

Given that dealer banks also share the same CARA utility as the client institutions, it is straightforward to follow the same steps as before to obtain their optimal asset demand as a solution to a mean-variance maximization problem given below:

$$\max_{\{q_{b,i},Q_{b,i},e_{b,i}\}} \Big[(\mu_{b,i}\pi^e - p)q_{b,i} + (\mu_{b,i}\Pi^e - P)Q_{b,i} + (\theta_d - s)e_{b,i}\Big] - \frac{\widetilde{\gamma}_{b,i}}{2}\Big[q_{b,i}^2\sigma_\pi^2 + Q_{b,i}^2\sigma_\Pi^2 + e_{b,i}^2\sigma_d^2$$
$$+ \sigma_{y_{b,i}}^2 + 2q_{b,i}Q_{b,i}\sigma_{\pi,\Pi} + 2q_{b,i}e_{b,i}\sigma_{\pi,d} + 2q_{b,i}\sigma_{\pi,y_{b,i}} + 2Q_{b,i}e_{b,i}\sigma_{\Pi,d} + 2Q_{b,i}\sigma_{\Pi,y_{b,i}}\Big]$$
$$\text{subject to:}\quad G_b^S(Q_{b,i}, z_{b,i}) \geq 0 \quad \text{and} \quad G_b^L(q_{b,i}, z_{b,i}) \geq 0 \tag{C.10}$$

Taking market prices as given, the first-order conditions necessary to attain a maximum are given by:

$$q_{b,i}^* = \frac{\mu_{b,i}\pi^e - p^*}{\widetilde{\gamma}_{b,i}\sigma_\pi^2} - \frac{\sigma_{\pi,\Pi}}{\sigma_\pi^2}Q_{b,i}^* - \frac{\sigma_{\pi,d}}{\sigma_\pi^2}e_{b,i}^* - \frac{\sigma_{\pi,y_{b,i}}}{\sigma_\pi^2} - \frac{\lambda_{b,i}^L g_b^L}{\widetilde{\gamma}_{b,i}\sigma_\pi^2} \tag{C.11}$$
$$Q_{b,i}^* = \frac{\mu_{b,i}\Pi^e - P^*}{\widetilde{\gamma}_{b,i}\sigma_\Pi^2} - \frac{\sigma_{\pi,\Pi}}{\sigma_\Pi^2}q_{b,i}^* - \frac{\sigma_{\Pi,d}}{\sigma_\Pi^2}e_{b,i}^* - \frac{\sigma_{\Pi,y_{b,i}}}{\sigma_\Pi^2} - \frac{\lambda_{b,i}^S g_b^S}{\widetilde{\gamma}_{b,i}\sigma_\Pi^2} \tag{C.12}$$
$$e_{b,i}^* = \frac{\theta_d - s^*}{\widetilde{\gamma}_{b,i}\sigma_d^2} - \frac{\sigma_{\pi,d}}{\sigma_d^2}q_{b,i}^* - \frac{\sigma_{\Pi,d}}{\sigma_d^2}Q_{b,i}^* \tag{C.13}$$

Next, we eliminate demand for the market asset $e^*_{b,i}$ from the system using substitution. After a bit of algebra, $q^*_{b,i}$ and $Q^*_{b,i}$ can be expressed respectively as:

$$q^*_{b,i} = \frac{\mu_{b,i}\pi^e - p^*}{\widetilde{\gamma}_{b,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} - \frac{\sigma_d}{\sigma_\pi}\left[\frac{\theta_d - s^*}{\widetilde{\gamma}_{b,i}\sigma^2_d(1-\rho^2_{\pi,d})}\right]\rho_{\pi,d} - \frac{\sigma_\Pi}{\sigma_\pi}\left[\frac{\rho_{\pi,\Pi} - \rho_{\pi,d}\rho_{\Pi,d}}{1-\rho^2_{\pi,d}}\right]Q^*_{b,i} \tag{C.14}$$

$$- \left[\frac{1}{(1-\rho^2_{\pi,d})\sigma^2_\pi}\right]\left(\sigma_{\pi,y_{b,i}} + \frac{\lambda^L_{b,i}g^L_b}{\widetilde{\gamma}_{b,i}}\right) \tag{C.15}$$

$$Q^*_{b,i} = \frac{\mu_{b,i}\Pi^e - P^*}{\widetilde{\gamma}_{b,i}\sigma^2_\Pi(1-\rho^2_{\Pi,d})} - \frac{\sigma_d}{\sigma_\Pi}\left[\frac{\theta_d - s^*}{\widetilde{\gamma}_{b,i}\sigma^2_d(1-\rho^2_{\Pi,d})}\right]\rho_{\Pi,d} - \frac{\sigma_\pi}{\sigma_\Pi}\left[\frac{\rho_{\Pi,\pi} - \rho_{\Pi,d}\rho_{\pi,d}}{1-\rho^2_{\Pi,d}}\right]q^*_{b,i} \tag{C.16}$$

$$- \left[\frac{1}{(1-\rho^2_{\Pi,d})\sigma^2_\Pi}\right]\left(\sigma_{\Pi,y_{b,i}} + \frac{\lambda^S_{b,i}g^S_b}{\widetilde{\gamma}_{b,i}}\right) \tag{C.17}$$

From the expressions above, it becomes clear that of $\rho_{\pi,\Pi} = \rho_{\pi,d}\rho_{\Pi,d}$ as we assumed, then $Q^*_{b,i}$ (and its exogenous shifters) are not part of the solution for $q^*_{b,i}$ and vice versa. The expression for asset demand from dealer banks in the main text can subsequently follows by using the definition for $\widetilde{\gamma}_{b,i} = \gamma_{b,i}/a_{b,i}$.

C.3 Proof of lemma 1 and lemma 2

The proofs follow directly from replacing the asset demand functions in propositions 1 and 2 into the market clearing conditions

$$\sum_{i\in\Theta_f}\widetilde{q}^*_{f,i} + \sum_{i\in\Theta_b}\widetilde{q}^*_{b,i} = 0 \tag{C.18}$$

$$\sum_{i\in\Theta_h}\widetilde{Q}^*_{h,i} + \sum_{i\in\Theta_b}\widetilde{Q}^*_{b,i} = 0 \tag{C.19}$$

and rearranging.

D Identification with granular instrumental variables

In this section, we elaborate further on the identification strategy using granular instrumental variables. For exposition, consider the case of identifying the demand shock in the long horizon inflation swap market in which case we have to find a granular IV for ε_f.

Recall the demand system for institution i among pension funds f, and append to it a time index t:

$$\frac{q^*_{f,i,t}}{a_{f,i,t}} = \frac{\mu_{f,i}\pi^e_t - p^*_t}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} - \left(\frac{\sigma_d}{\sigma_\pi}\right)\left[\frac{\theta_{d,t} - s^*_t}{\gamma_{f,i}\sigma^2_d(1-\rho^2_{\pi,d})}\right]\rho_{\pi,d} \underbrace{-\left[\frac{1}{(1-\rho^2_{\pi,d})\sigma^2_\pi}\right]\left(\frac{\sigma_{\pi,y_{f,i,t}}}{a_{f,i,t}} + \frac{\lambda_{f,i,t}g_{f,i,t}}{\gamma_{f,i}}\right)}_{=\varepsilon_{f,i,t}} \tag{D.1}$$

The fund-specific demand shock is denoted by $\varepsilon_{f,i,t}$.

Recall also the combined results in lemmas 1 and 2 that solve for the observed market price for the swap contract in the long market as a combination of expected inflation minus a compensation for risk premia (the fundamental component) and liquidity premia:

$$p^*_t = \Lambda\pi^e_t - rp^*_t + lp^*_t \tag{D.2}$$

where the components are defined as:

$$\Lambda = \frac{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t}\mu_{f,i}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} + \frac{\sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}\mu_{b,i}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} \tag{D.3}$$

$$rp^*_t = \left(\frac{\theta_{d,t} - s^*_t}{\sigma^2_d}\right)\sigma_{\pi,d} \tag{D.4}$$

$$lp^*_t = -\frac{\sum_{i\in\Theta_b}\left(\sigma_{\pi,y_{b,i}} + \frac{\lambda^L_{b,i,t}g^L_{b,i,t}}{\widetilde{\gamma}_{b,i,t}}\right)}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} - \frac{\sum_{i\in\Theta_f}\left(\sigma_{\pi,y_{f,i}} + \frac{\lambda_{f,i,t}g_{f,i,t}}{\widetilde{\gamma}_{f,i,t}}\right)}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} \tag{D.5}$$

Finally, recall also the definition of $\tilde{\varepsilon}_{f,i,t}$:

$$\tilde{\varepsilon}_{f,i,t} = \varepsilon_{f,i,t} - \frac{\kappa^{lp}_{f,i}lp^*_t}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} \tag{D.6}$$

In the equation above, $\tilde{\varepsilon}_{f,i,t}$ refers to the idiosyncratic component of the fund-specific demand shock $\varepsilon_{f,i,t}$ and $\kappa^{lp}_{f,i}$ captures the contribution of pension fund i's liquidity frictions to the market-wide liquidity premium lp^*_t that was defined in Equation (D.5).

Substituting (D.2), (D.4) and (D.6) into Equation (D.1) and rearranging terms, the demand system above can be rewritten as:

$$\frac{q^*_{f,i,t}}{a_{f,i,t}} = \frac{(\mu_{f,i}-\Lambda)\pi^e_t + (\kappa^{lp}_{f,i}-1)lp^*_t}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} + \widetilde{\varepsilon}_{f,i,t} = \boldsymbol{\omega}'_{f,i}\mathbf{F}_t + \widetilde{\varepsilon}_{f,i,t} \tag{D.7}$$

where $\mathbf{F}_t$ and $\boldsymbol{\omega}_{f,i}$ are respectively the common factors and the fund-specific factor loadings:

$$\mathbf{F}_t = \begin{pmatrix}\pi^e_t \\ lp^*_t\end{pmatrix}, \quad \boldsymbol{\omega}_{f,i} = \begin{pmatrix}\frac{(\mu_{f,i}-\Lambda)}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} \\ \frac{\kappa^{lp}_{f,i}-1}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})}\end{pmatrix} \tag{D.8}$$

This separates idiosyncratic shocks to institutional demand from the common factors, $\mathbf{F}_t$.

Our empirical implementation estimates a modified regression equation that allows for persistence in institutional demand and for fund-specific and time fixed effects:

$$\frac{q_{f,i,t}}{a_{f,i,t}} = \alpha_{f,i} + \tau_t + \sum_{j=1}^{J} \beta_j \frac{q_{f,i,t-j}}{a_{f,i,t-j}} + \boldsymbol{\omega}'_{f,i}\mathbf{F}_t + \widetilde{\varepsilon}_{f,i,t}. \tag{D.9}$$

We use $J = 3$ lags to be consistent with the lags used in estimation of the Bayesian sign restrictions SVAR.

Although our model implies a two-factor model, we allow for a larger number of factors to capture other sources of heterogeneity within the sector orthogonal to the model's components that could lead to disturbances in demand in the data, such as differences in fund structures (e.g., liability-driven investment, defined contributions). The number of factors are first chosen according to Bai and Ng (2002), and we adjust them appropriately such that the structural responses of the demand and supply shocks satisfy the sign restrictions imposed. This procedure led to 21 factors being estimated for PFLDI demand, 10 for hedge fund demand and 10 for dealer bank supply.

The granular IV for pension fund demand that emerges from this procedure is given by:

$$GIV_{f,t} = \sum_{i\in\Theta_f} a_{f,i,t}\widetilde{\varepsilon}_{f,i,t} \tag{D.10}$$

Recall from Equation (12) that the sector-wide demand shock from pension funds $\varepsilon_{f,t}$ is given by:

$$\varepsilon_{f,t} = -\frac{\sum_{i\in\Theta_f}\left\{\sigma_{\pi,y_{f,i}} + \frac{\lambda_{f,i}g_{f,i,t}}{\widetilde{\gamma}_{f,i,t}}\right\}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} \tag{D.11}$$

Given the definition of $\varepsilon_{f,i,t}$ in Equation (D.6), the equation above can be rewritten as:

$$\varepsilon_{f,t} = \frac{\sum_{i\in\Theta_f}(\kappa^{lp}_{f,i}lp^*_t)\widetilde{\gamma}^{-1}_{f,i,t}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} + (1-\rho^2_{\pi,d})\sigma^2_\pi \frac{GIV_{f,t}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} \tag{D.12}$$

It is straightforward to see that $GIV_{f,t}$ qualifies as an instrumental variable for $\varepsilon_{f,t}$. Insofar as there is some granularity in $a_{f,i,t}$, $GIV_{f,t}$ will satisfy the relevance condition since $\mathbb{E}(GIV_{f,t}\varepsilon_{f,t}) \neq 0$.[58] For exclusion restriction, note that $\tilde{\varepsilon}_{fi,t} \perp \varepsilon_{x,t}, \varepsilon_{b,t}, \varepsilon_{\pi,t}$ by construction, since these three shocks are spanned by $F_{f,t}$. Hence, $\mathbb{E}(GIV_{f,t}\varepsilon_{x,t}) = 0$, $\mathbb{E}(GIV_{f,t}\varepsilon_{b,t}) = 0$ and $\mathbb{E}(GIV_{f,t}\varepsilon_{\pi,t}) = 0$. Thus, $GIV_{f,t}$ satisfies all required moments to qualify as a valid instrument for $\varepsilon_{f,t}$.

We subsequently apply the same procedure on data for hedge funds and dealer banks to obtain $GIV_{x,t}$ and $GIV_{b,t}$ as instruments for $\varepsilon_{x,t}$ and $\varepsilon_{b,t}$, which are the demand shocks in the short-horizon market and shocks to dealers' supply functions. We then use the simultaneous equations from the demand system to

[58] No granularity means $GIV_{f,t} = 0$ as all the idiosyncratic shocks average out.

back out the granular IV for the fundamental shock. Recall that its static representation is given by:

$$\begin{pmatrix} Q_t \\ P_t \\ q_t \\ p_t \end{pmatrix} = constant + \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix} \underbrace{\begin{pmatrix} \varepsilon_{x,t} \\ \varepsilon_{f,t} \\ \varepsilon_{b,t} \\ \varepsilon_{\pi,t} \end{pmatrix}}_{\mathbf{B}} = constant + \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \\ u_{4,t} \end{pmatrix} \tag{D.13}$$

Where each $\mathbf{b}_i$ for $i \in \{1, 2, 3, 4\}$ is a 4×1 column vector. We then use $GIV_{f,t}$, $GIV_{x,t}$ and $GIV_{b,t}$ as instruments to project $u_{4,t}$ on $u_{1,t}, u_{2,t}, u_{3,t}$. The residual that emerges from this procedure is then a valid instrument for $\varepsilon_{\pi,t}$, which we label as $GIV_{\pi,t}$. To see this, note that the matrix B by assumption can be inverted to obtain:

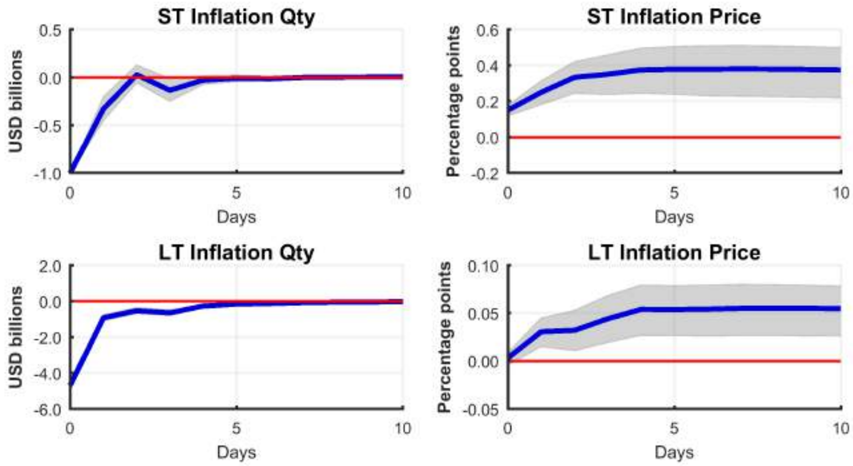
$$u_{4,t} = a_{4,1}u_{1,t} + a_{4,2}u_{2,t} + a_{4,3}u_{3,t} + \varepsilon_{\pi,t} \tag{D.14}$$

The residuals from the 2SLS regression exactly yields $\varepsilon_{\pi,t}$.[59] Next, we project u_t sequentially on $GIV_{f,t}$, $GIV_{x,t}$, $GIV_{b,t}$ and $IV_{\pi,t}$. This identifies the coefficients of the structural impact matrix b_1 to b_4 up to sign and scale.

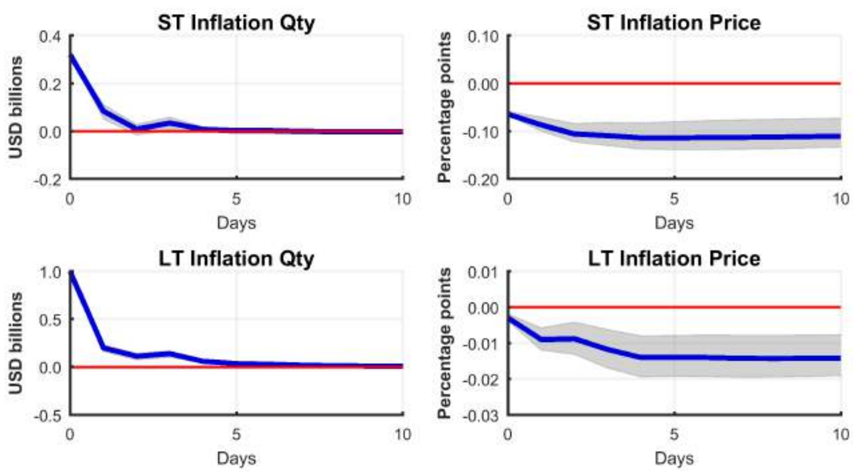
[59]The numbering of the u's is arbitrary. One could also project $u_{1,t}$ on $u_{2,t}$, $u_{3,t}$ and $u_{4,t}$ instead and get a different $IV_{\pi,t}$ but this would be perfectly collinear with the instrument arising from the alternative projection.

Figure D.1 ESTIMATED IMPULSE RESPONSE FUNCTIONS TO LIQUIDITY SHOCKS FROM A GIV STRATEGY

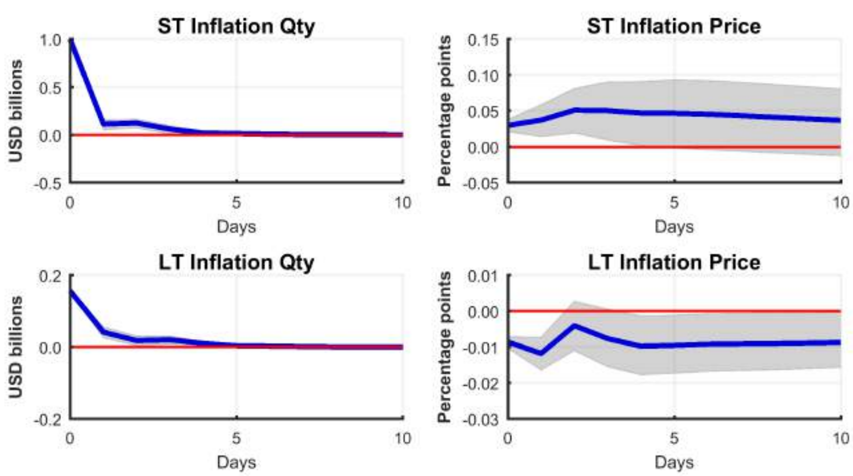
(a) Dealer supply shock ($\varepsilon_{b,t}$)



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



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



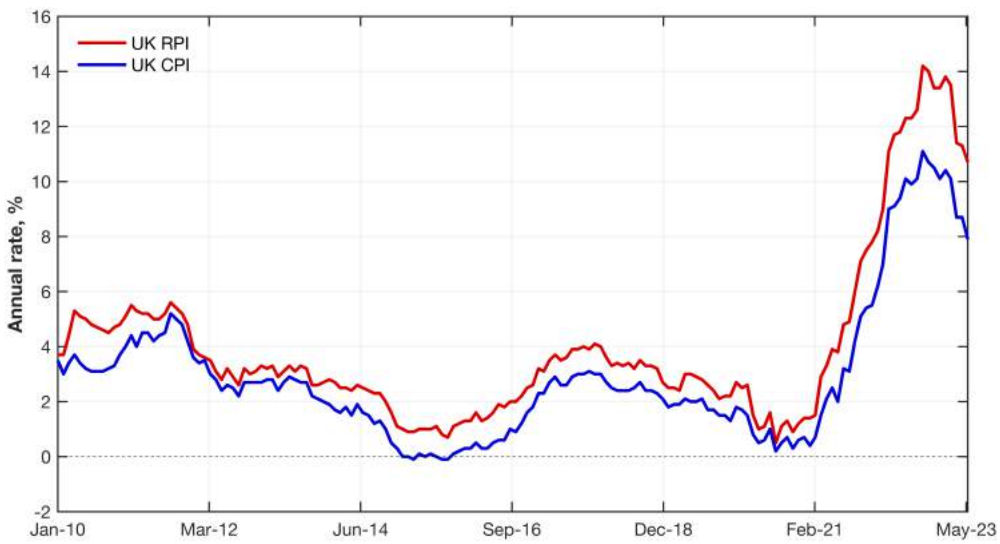
E Additional Figures

This section includes additional empirical results that were mentioned in the main text.

E.1 The RPI-CPI wedge

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

Figure E.1 Comparison between UK RPI and UK CPI



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

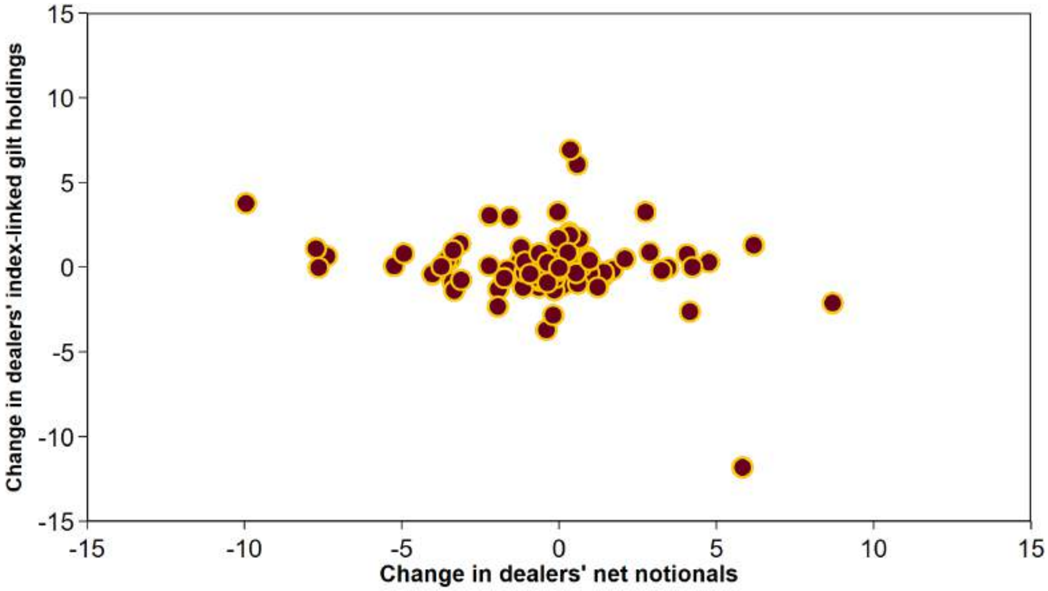
E.2 Dealers' swap net notionals vs. index-linked gilt holdings

Figure E.2 shows that the quarterly change in the holdings of index-linked government bonds of individual dealer banks is only weakly correlated with the change in their net notionals in UK RPI inflation swaps (with a correlation coefficient of -0.23).

E.3 Pension funds' index-linked gilt holdings

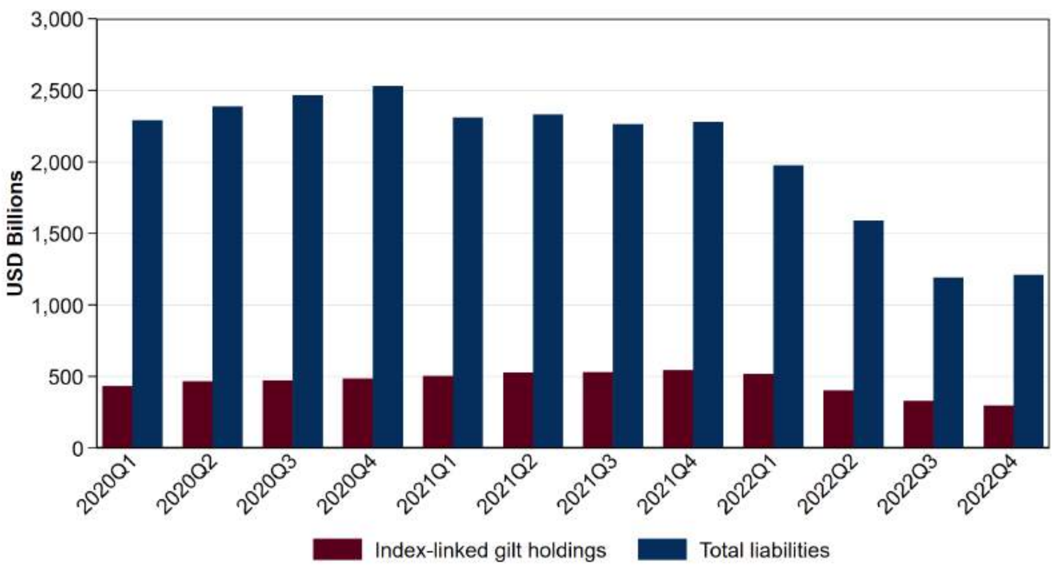
Figure E.3 shows that the holdings of index-linked government bonds by pension funds is well below their total liabilities. Arguably, these holdings to not exhaust the funds' need for inflation protection, leading them to turn to the inflation swap market.

Figure E.2 DEALERS' INDEX-LINKED GILT HOLDINGS VERSUS SWAP NET NOTIONALS



NOTE: Comparison of the quarterly change in index-linked gilt holdings of individual dealer banks and the quarterly change in their net notionals in UK RPI inflation swaps in $ billions. SOURCE: UK banking system's Global Network of granular exposures and DTCC Trade Repository OTC interest rate trade state files, from March 2019 to June 2022.

Figure E.3 PENSION FUNDS' INDEX-LINKED GILT HOLDINGS VERSUS TOTAL LIABILITIES

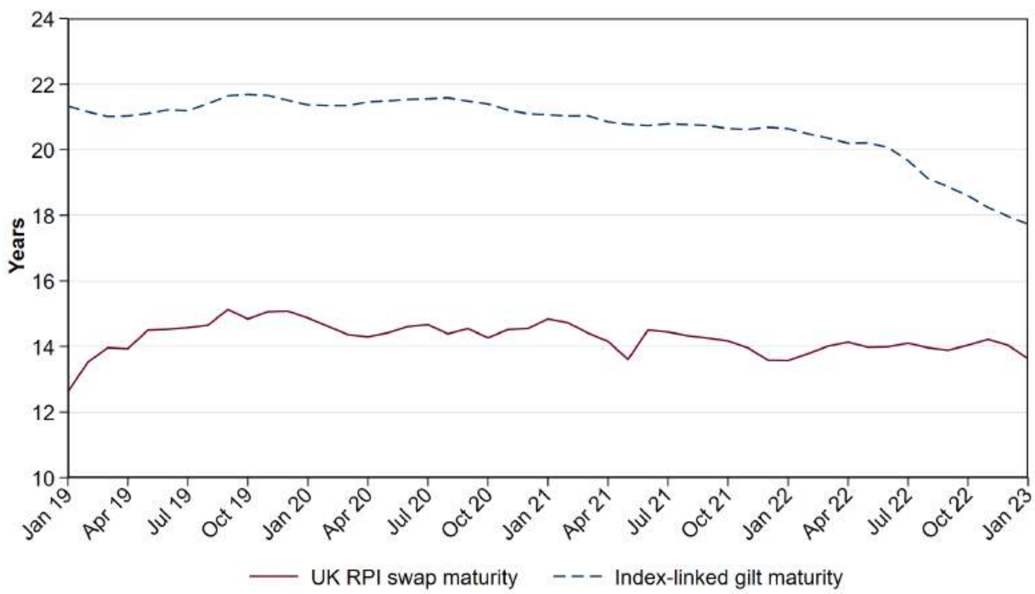


NOTE: Comparison of pension funds' index-linked gilt holdings and total liabilities in $ billions. SOURCE: Pension Protection Fund PPF 7800 Data and Office for National Statistics.

E.4 Average maturity of pension funds' UK RPI swaps and index-linked gilts

Figure E.4 shows that the average maturity of index-linked government bonds held by pension funds, approximated by pension funds' stock of index-linked gilts used as collateral in the gilt repo market, is around six to seven years longer the maturity of their outstanding UK RPI swap contracts. This is likely due to the fact that the UK Debt Management Office (DMO) tends to issue very long-dated index-linked gilts, often with a maturity of more than 30 years. Pension funds might find it difficult to negotiate similar, extremely long maturities with dealer banks in the UK RPI swap market.

Figure E.4 Average maturity of pension funds' UK RPI swaps and index-linked gilts



Note: Comparison of the maturity of index-linked government bonds held by pension funds, approximated by pension funds' stock of index-linked gilts used as collateral in the gilt repo market, and the maturity of pension funds' outstanding UK RPI swap contracts in years. Both maturities are volume-weighted averages. Source: Sterling Money Market Database and DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

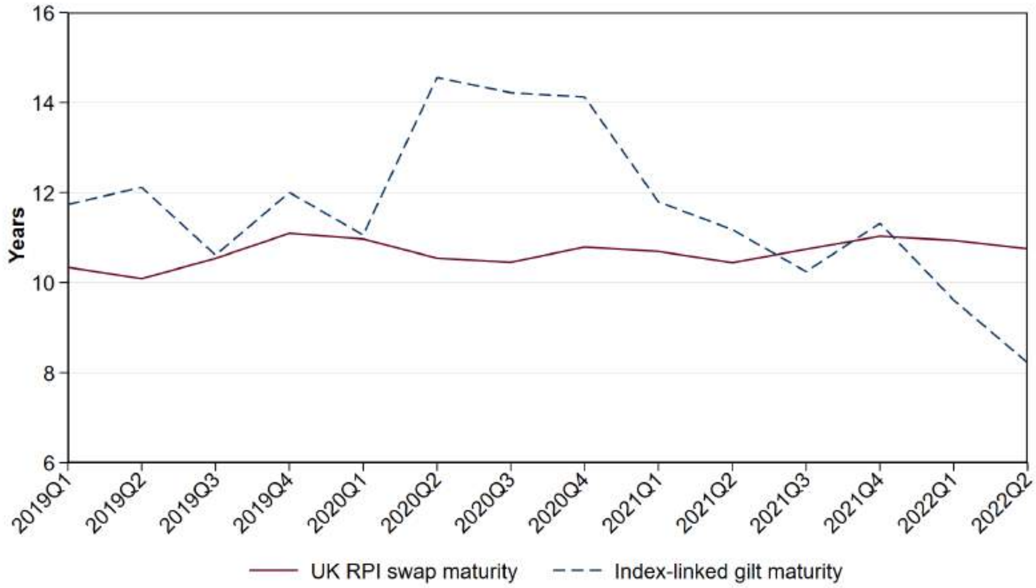
E.5 Average maturity of dealers' UK RPI swaps and index-linked gilts

Figure E.5 shows that the average maturity of index-linked government bonds held by dealer banks is well aligned with the maturity of their outstanding UK RPI swap contracts at the beginning of our sample period. In the second half of our sample, the two average maturities start to deviate, but the average swap maturity remains remarkably stable at around ten years.

E.6 Gross notional positions in the DTCC trade repository

Figure E.6 decomposes the inflation swaps in our dataset by the inflation index that they contract on. Recall that our data captures all trading where one side of the trade is based in the United Kingdom. Unsurprisingly, UK inflation swaps dominate, although there is a non-negligible share of inflation swaps

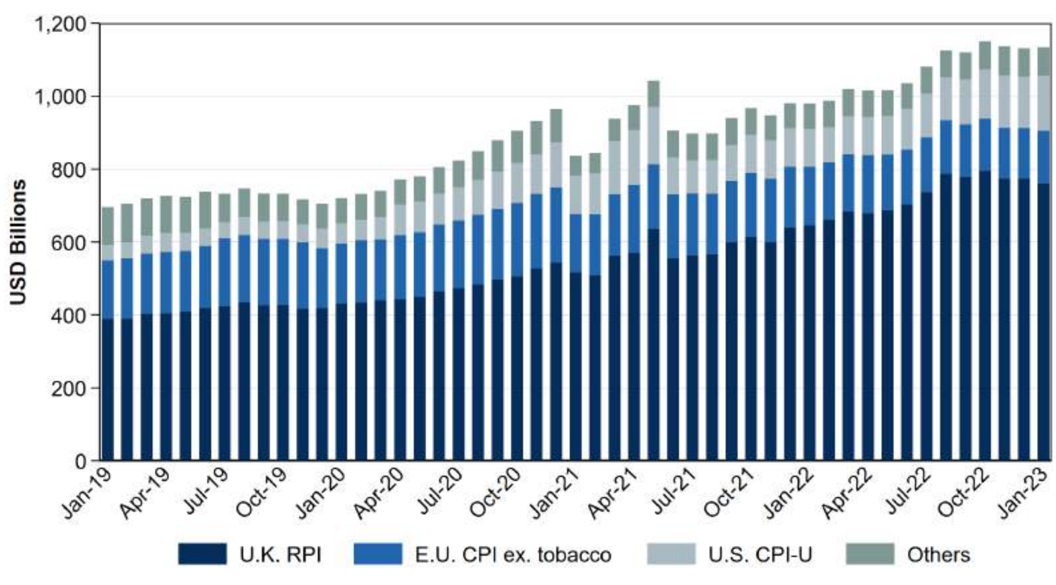
Figure E.5 Average maturity of dealers' UK RPI swaps and index-linked gilts



Note: Comparison of the maturity of index-linked government bonds held by dealers and the maturity of dealers' outstanding UK RPI swap contracts in years. Both maturities are volume-weighted averages. Source: UK banking system's Global Network of granular exposures and DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

in the data indexed to inflation in the EU and the US. WE focus on the inflation swaps that refer to the UK RPI.

Figure E.6 Outstanding positions by underlying inflation index



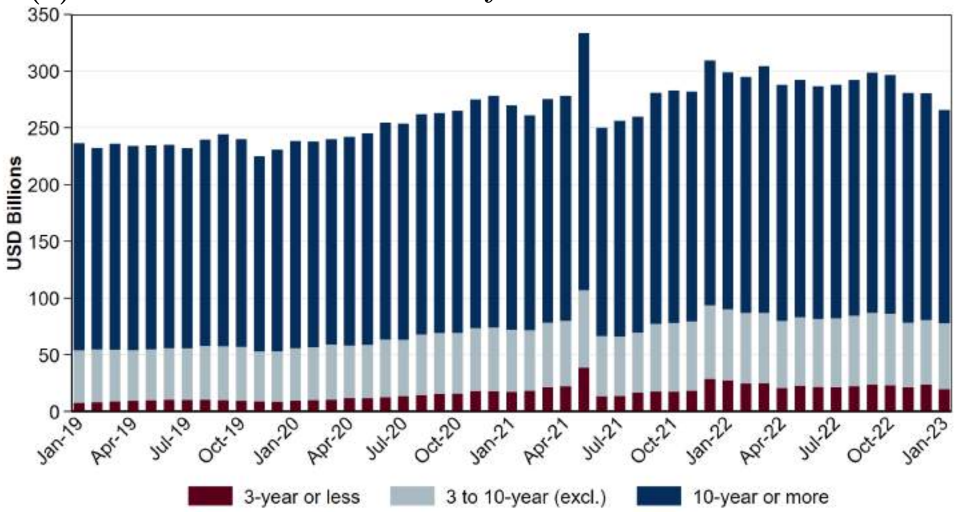
Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

E.7 Gross notional positions in the UK RPI inflation swap market

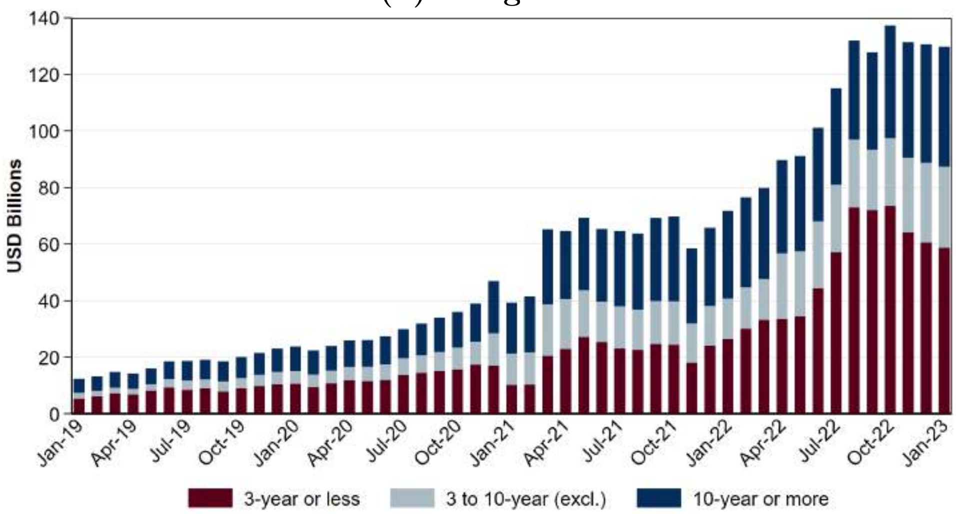
Figure E.7 reveals the gross notional holdings of the three sectors in our analysis by maturity across time. The segmentation of the market is clear also in gross holdings, even though it is starker on net holdings. Figure E.8 shows the gross notional positions of all other investor types in the market.

Figure E.7 GROSS NOTIONAL POSITIONS: THE MAIN AGENTS IN THE UK RPI MARKET

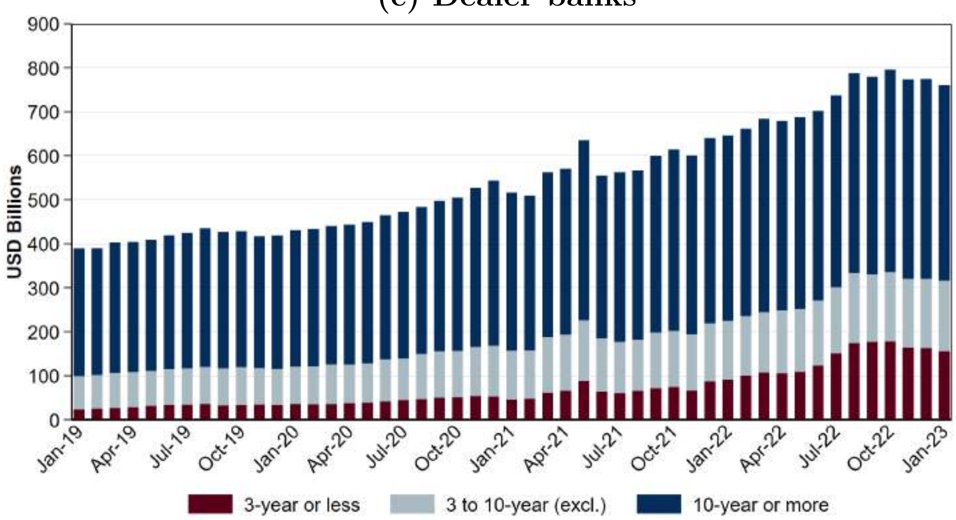
(a) Pension funds and liability investment driven institutions



(b) Hedge funds

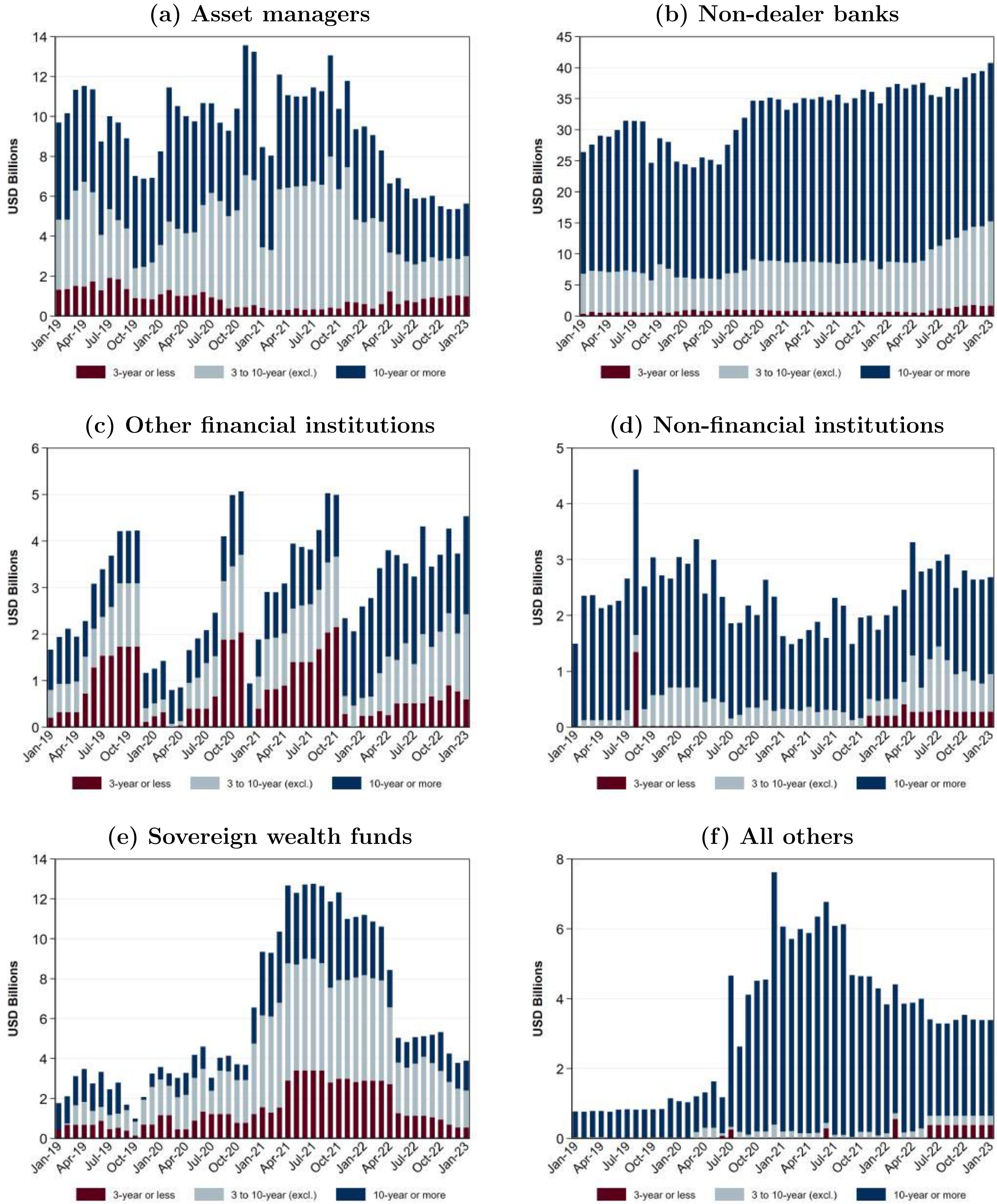


(c) Dealer banks



SOURCE: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

Figure E.8 Gross notional positions: The remainder of the market

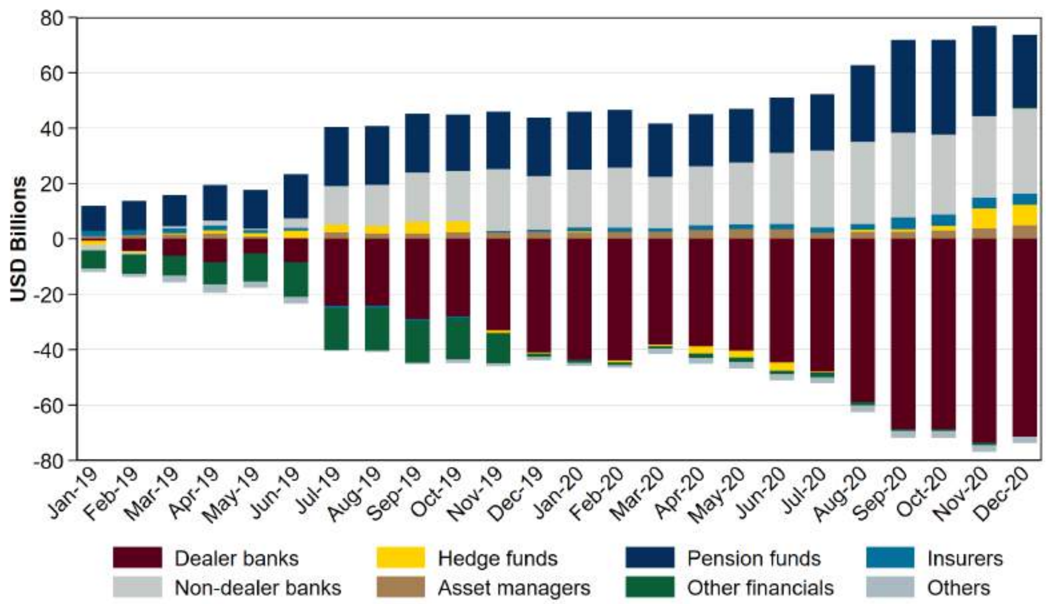


Note: "All others" include: state, supranational, proprietary trading firms, trading services and central banks. Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

E.8 Net notional positions in the UK RPI inflation market: all classifications

E.9 Net notional positions in the EU inflation market

Figure E.10 Dealers have a non-zero net exposure to inflation risk

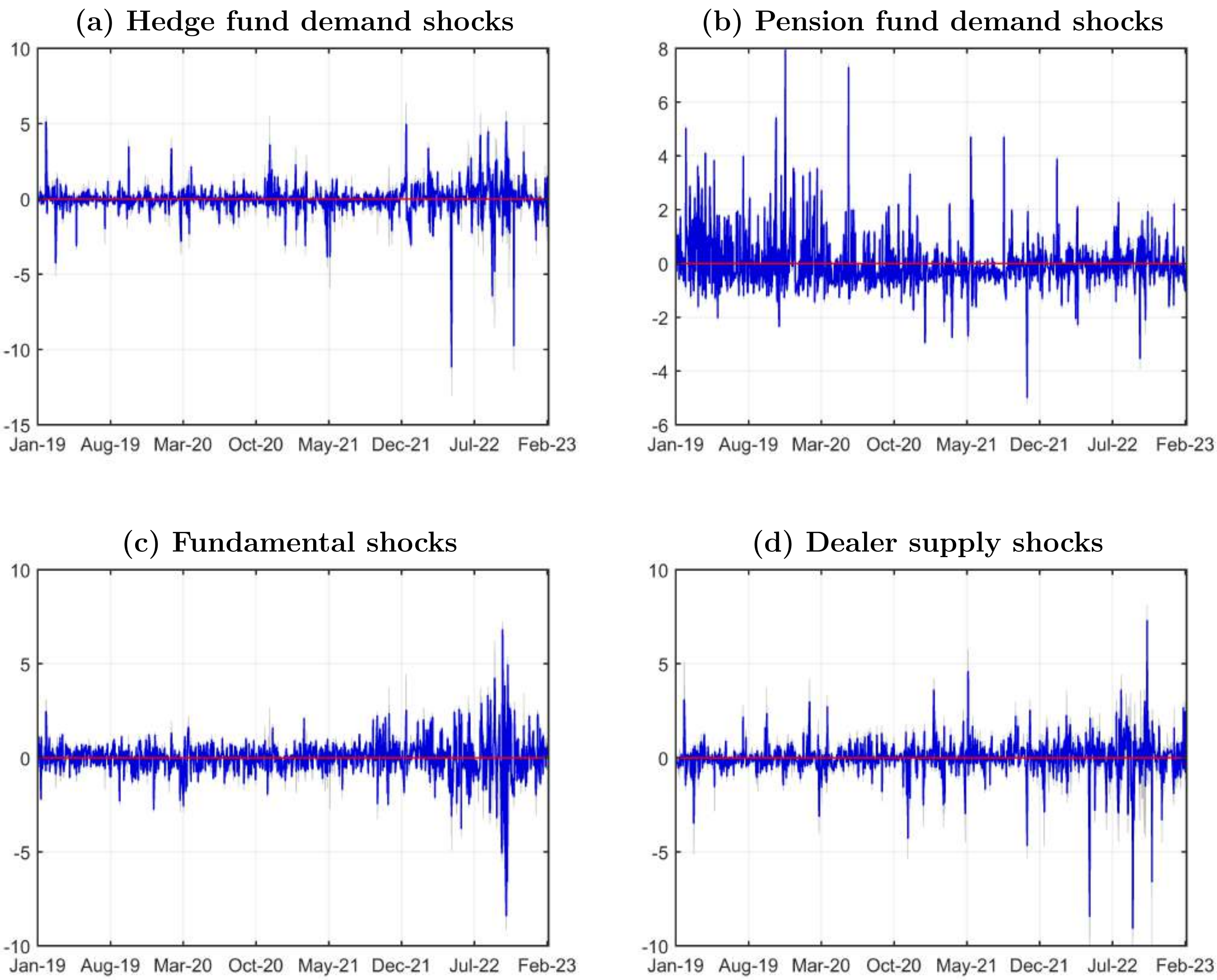


Note: Net notional positions in the EU inflation swap market only include the trading of inflation swaps written on the EU HICP index and the EU CPI index. We do not include in this calculation the positions traded on EU member country-specific inflation indexes. Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

E.10 Estimated structural shocks

Figures E.11 plots the time series of the four shocks identified by the sign restrictions strategy that follow from the differential responsiveness assumption 3.

Figure E.11 Estimated fundamental and liquidity shocks

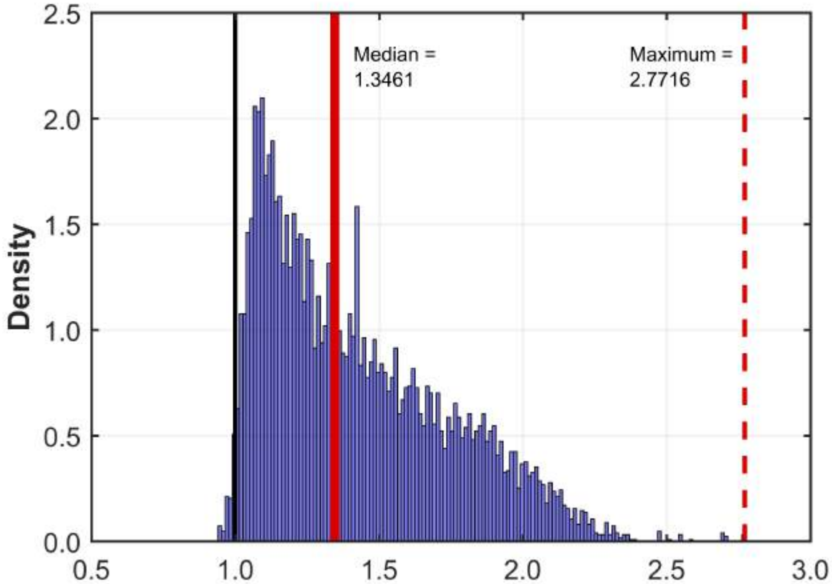


E.11 Cross-verification of identification strategies

Figure E.12 uses the identified shocks according to the sign restrictions to calculate their variance at the dates of RPI releases as well as at all other dates. The histogram of the ratio of the variances per draw form the posterior is plotted. The distribution points to strong evidence of the ratio being above 1, which is the identification assumption 5 behind the heteroskedasticity strategy.

Figure E.13 uses the identified shocks according to the sign restrictions to calculate the sample analog of the orthogonality conditions behind the granular instrument variable strategy. The figures shows the histograms per draw from the posterior distribution. They are all quite close to zero, validating assumption 4.

Figure E.12 Posterior distribution of ratio of variances



E.12 Desk separation at high-frequencies

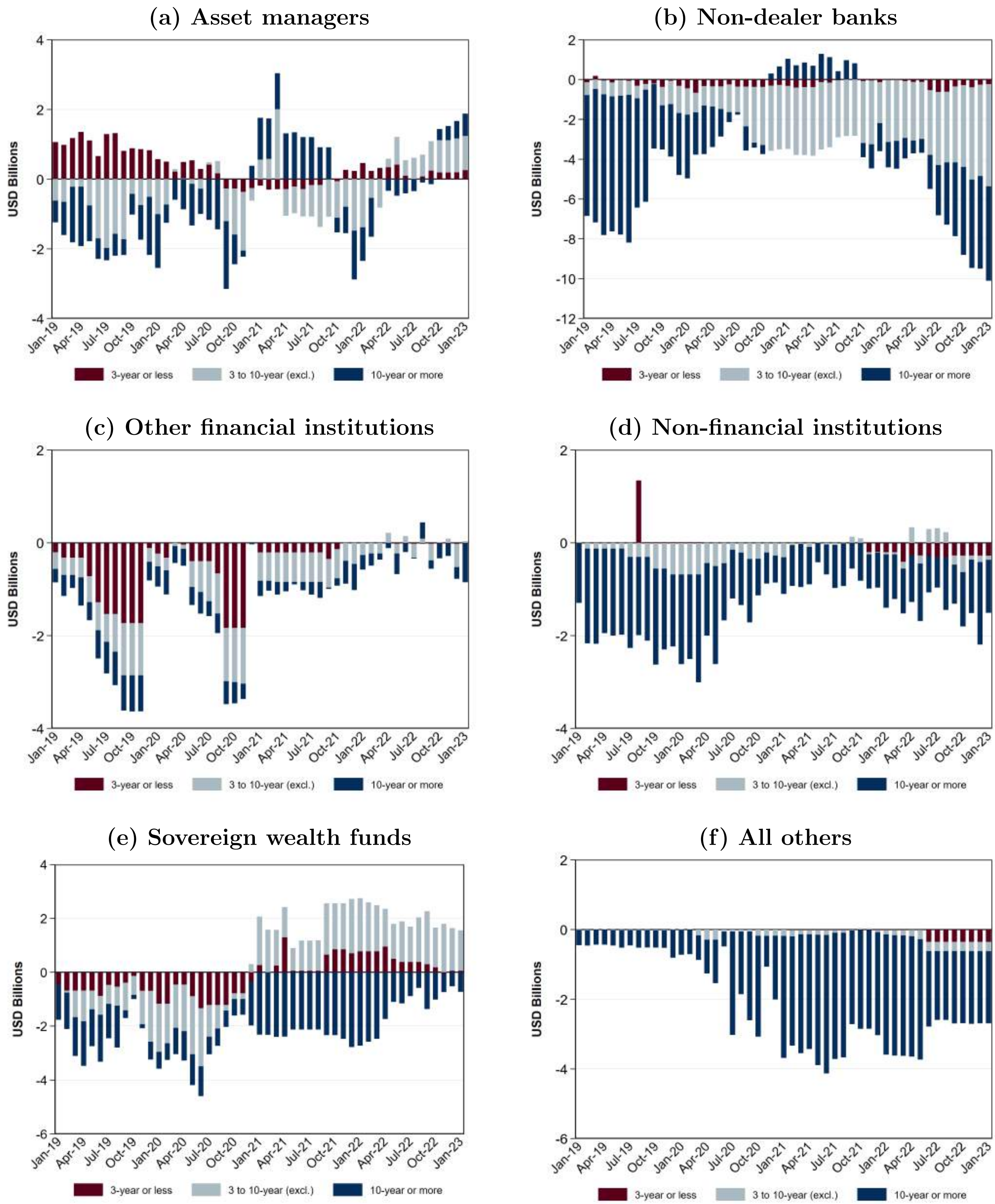
We made two assumptions when identifying shocks under a sign-restricted approach. In this section we provide further justification for making the desk separation assumption, which enabled us to identify the structural responses to demand shocks in both the short horizon and long horizon markets. Under this 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 E.14 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.

E.13 Historical decompositions

Figures E.15 and E.16 show the historical decompositions of the inflation swap prices into the contributions from each of the four shocks, for the long-horizon and short-horizon markets, respectively.

E.14 Disagreement about inflation: Evidence from trading activities

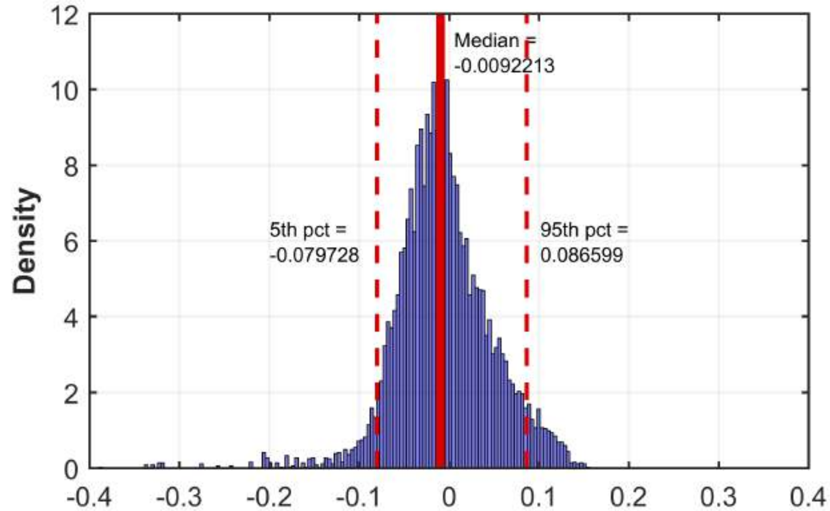
Figure E.9 NET NOTIONAL POSITIONS: THE REMAINDER OF THE MARKET



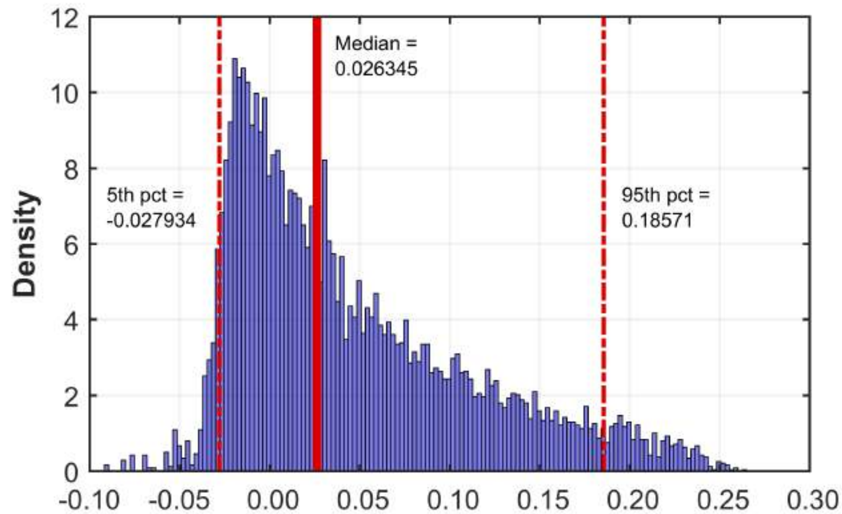
NOTE: "All others" include: state, supranational, proprietary trading firms, trading services and central banks. SOURCE: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

Figure E.13 Posterior distribution of second moments

(a) Sample correlation between $GIV_{f,t}$ and $\varepsilon_{\pi,t}$



(b) Sample correlation between $GIV_{h,t}$ and $\varepsilon_{\pi,t}$



(c) Sample correlation between $GIV_{b,t}$ and $\varepsilon_{\pi,t}$

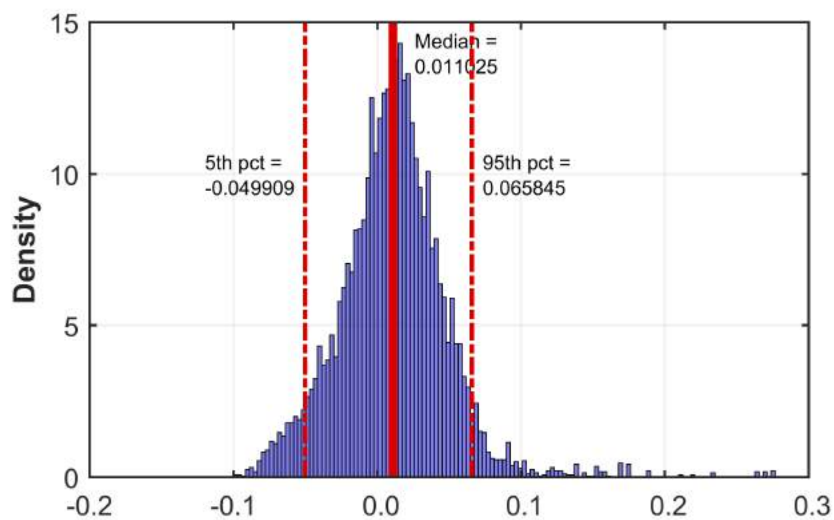
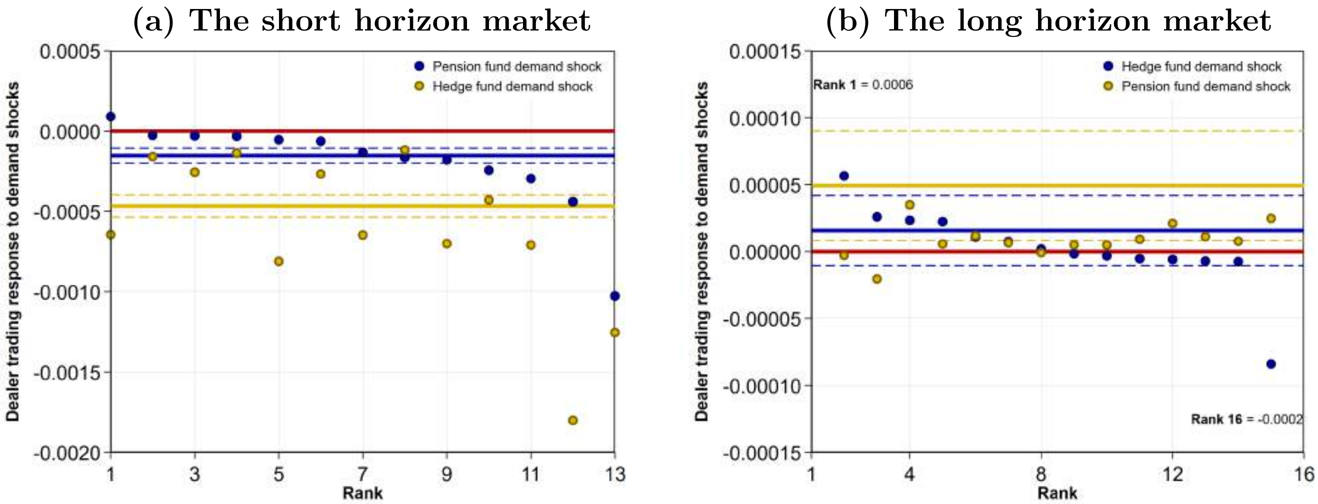


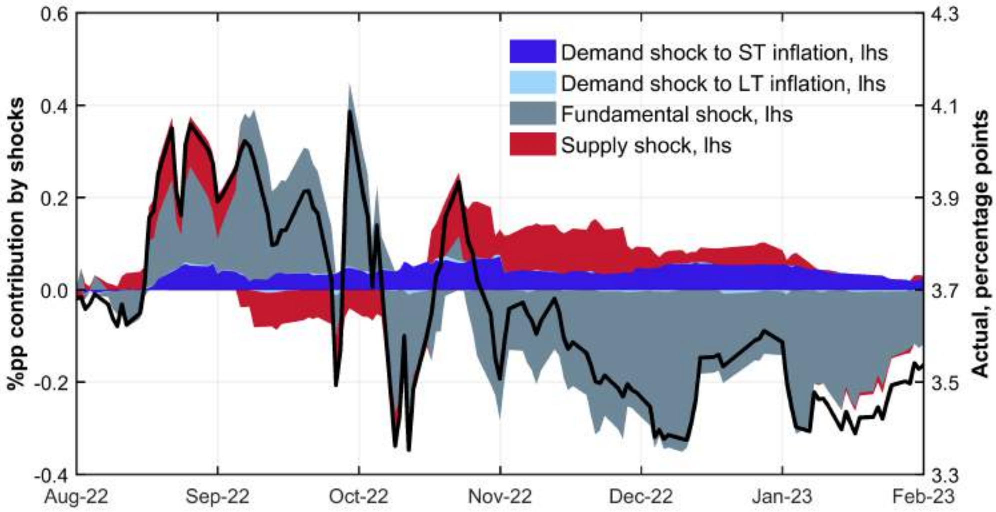
Figure E.14 Test of the desk separation assumption



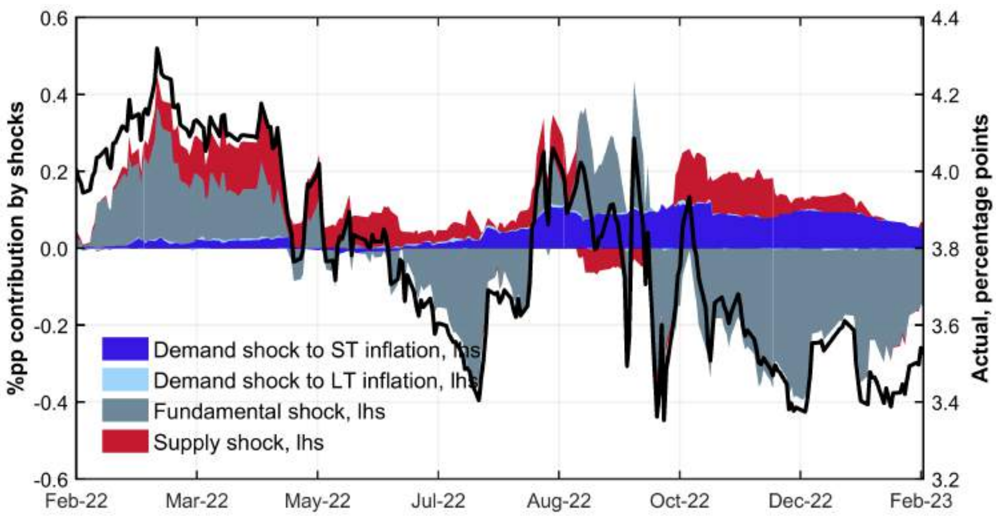
Note: The thick blue lines are pooled coefficient estimates with standard errors clustered at the institutional-level (where each is identified by a legal entity identifier), dashed blue lines are their 95% confidence intervals. Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

Figure E.15 Historical decomposition: long horizon inflation swap rates

(a) From August 2022 (LDI crisis period)



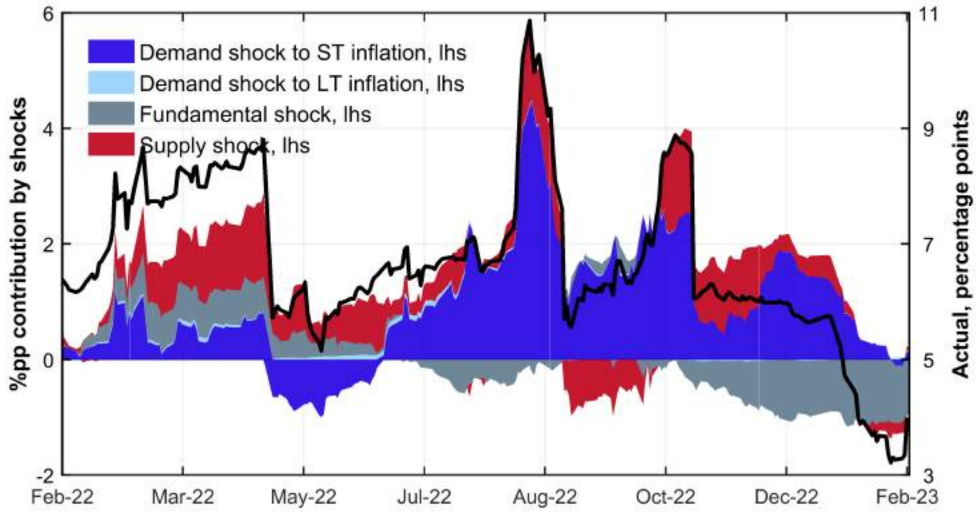
(b) From February 2022 (Ukraine War period)



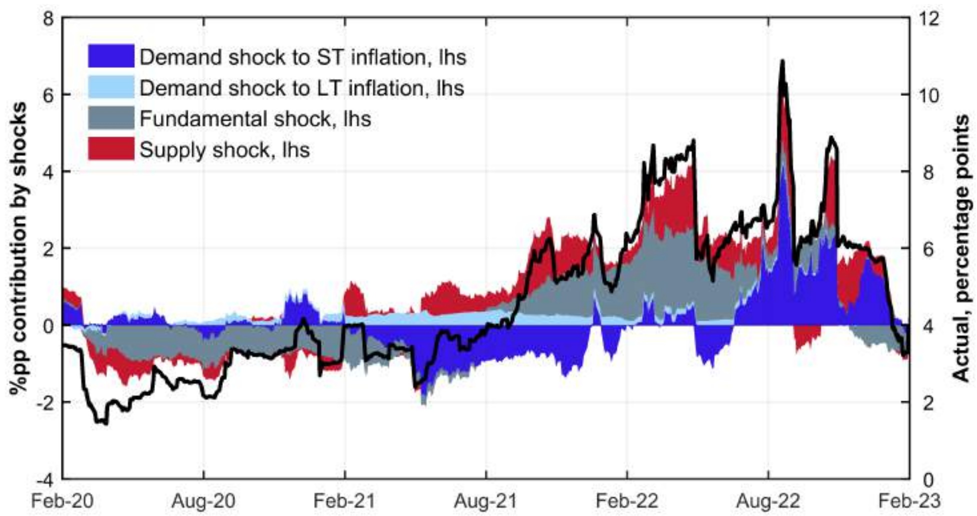
(c) From February 2020 (COVID-19 period)

Figure E.16 Historical decomposition: short horizon inflation swap rates

(a) From February 2022 (Ukraine War period)



(b) From February 2020 (Covid-19 period)



(c) Full sample

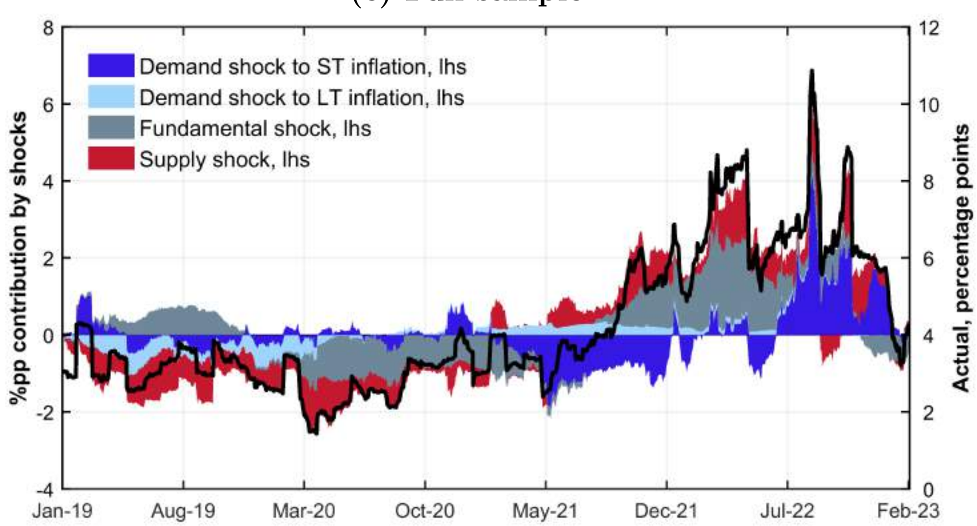
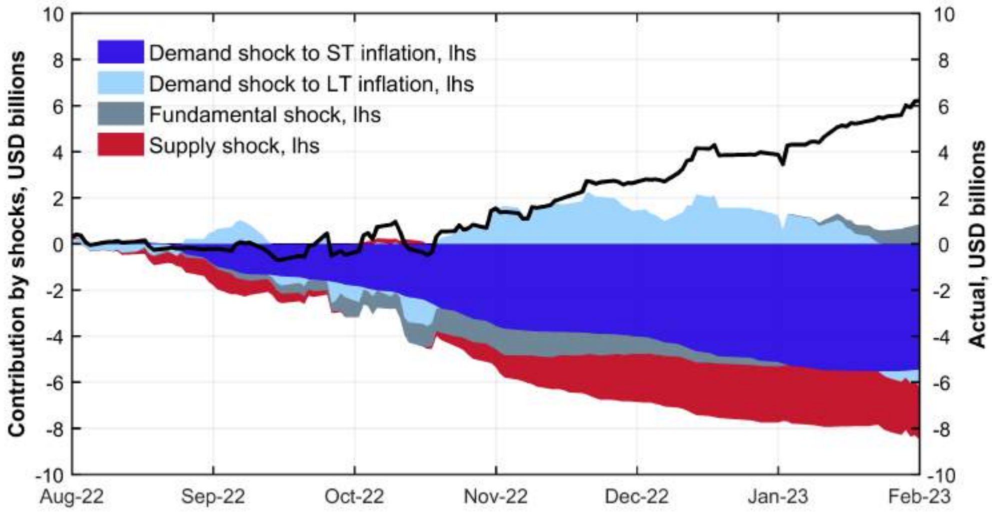
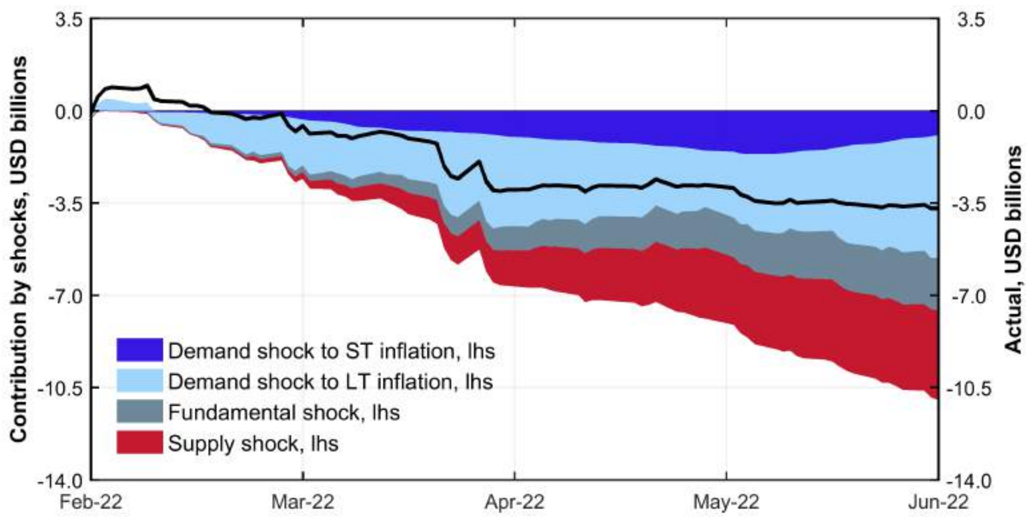


Figure E.17 Historical decomposition: net cumulative purchases of long horizon inflation swaps

(a) From August 2022 (LDI crisis period)



(b) From February 2022 (Ukraine War period)



(c) From February 2020 (COVID-19 period)

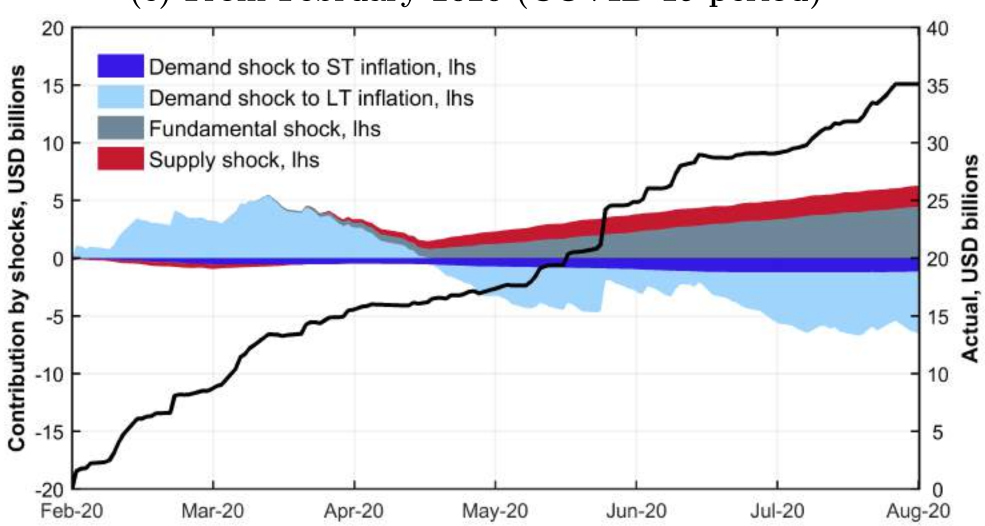
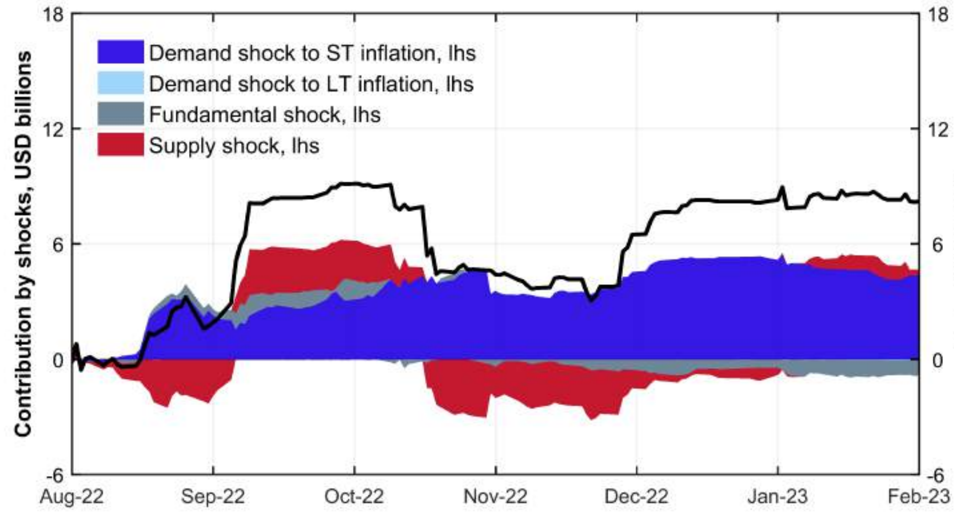
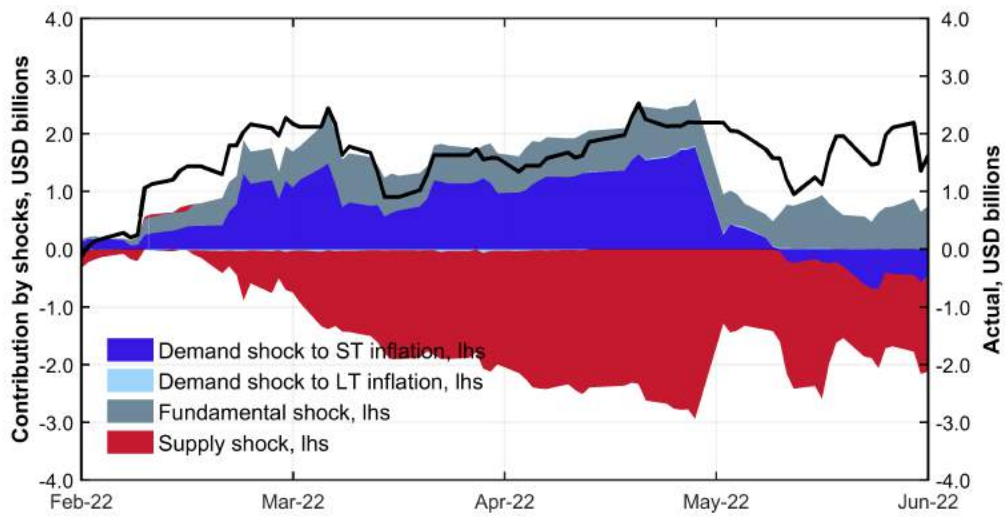


Figure E.18 Historical decomposition: net cumulative purchases of short horizon inflation swaps

(a) From August 2022 (LDI crisis period)



(b) From February 2022 (Ukraine War period)



(c) From February 2020 (COVID-19 period)

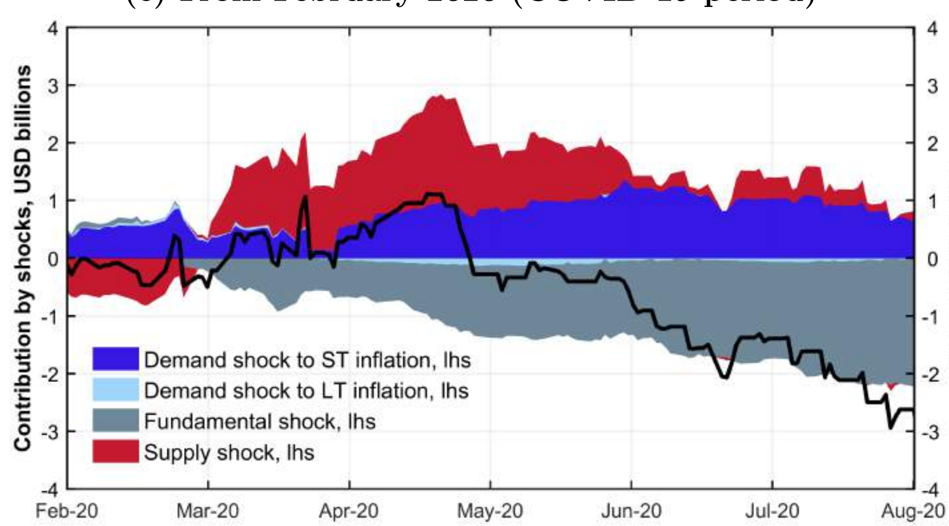
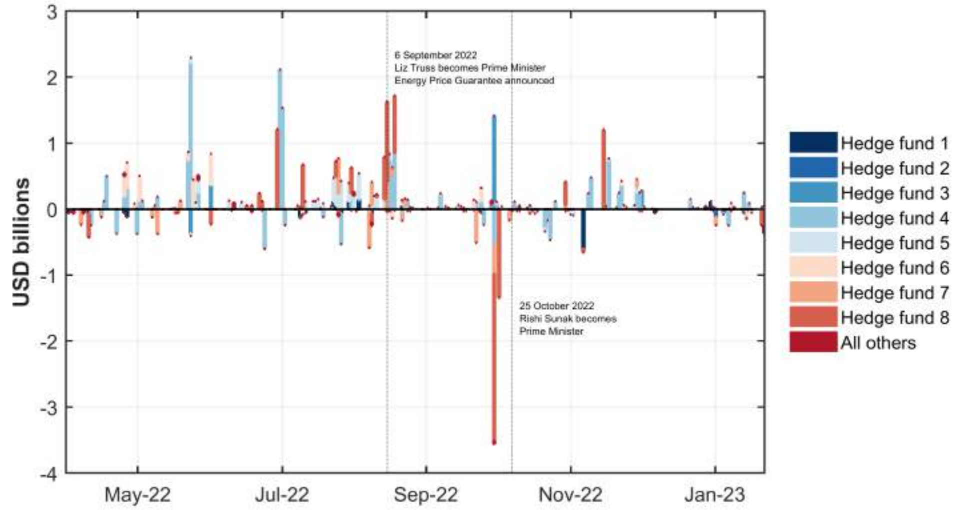
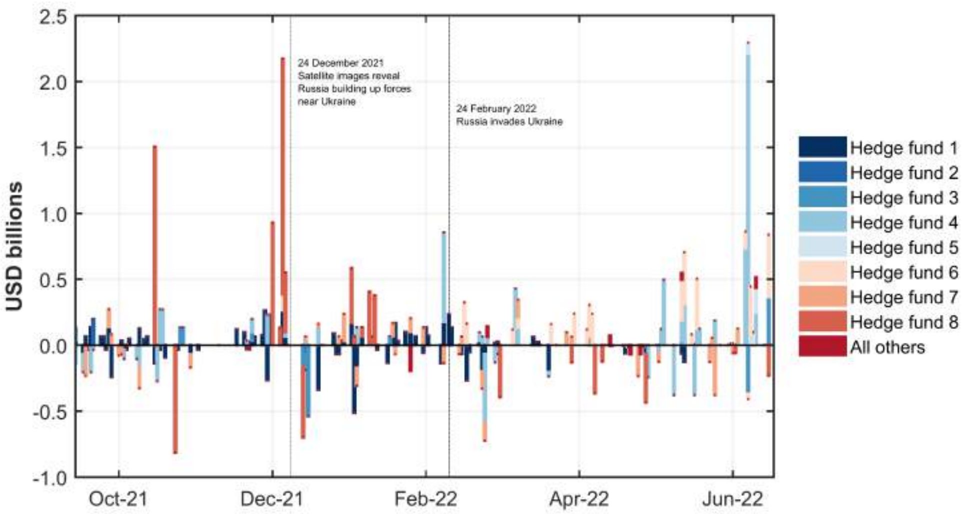


Figure E.19 Disagreement amongst hedge funds in the short horizon market

(a) From August 2022 (LDI crisis period)



(b) From February 2022 (Ukraine War period)



(c) From February 2020 (COVID-19 period)

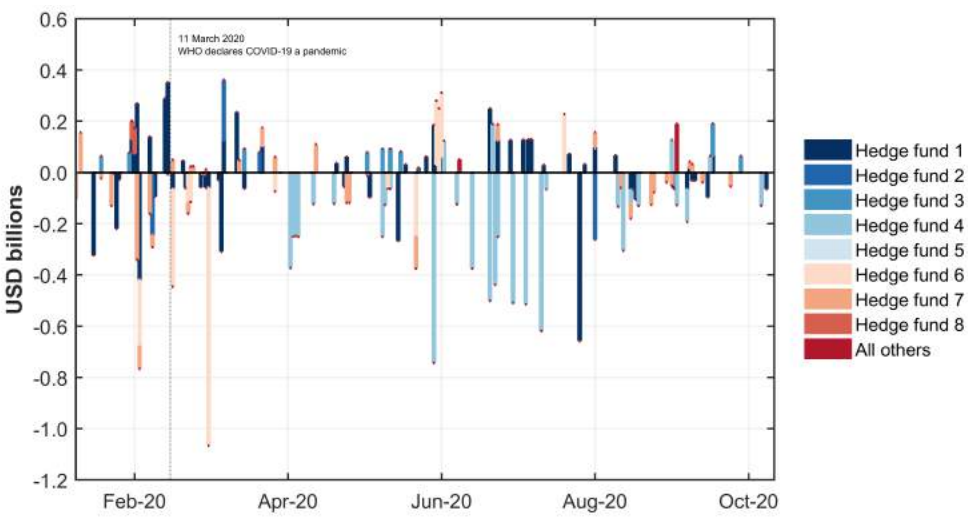
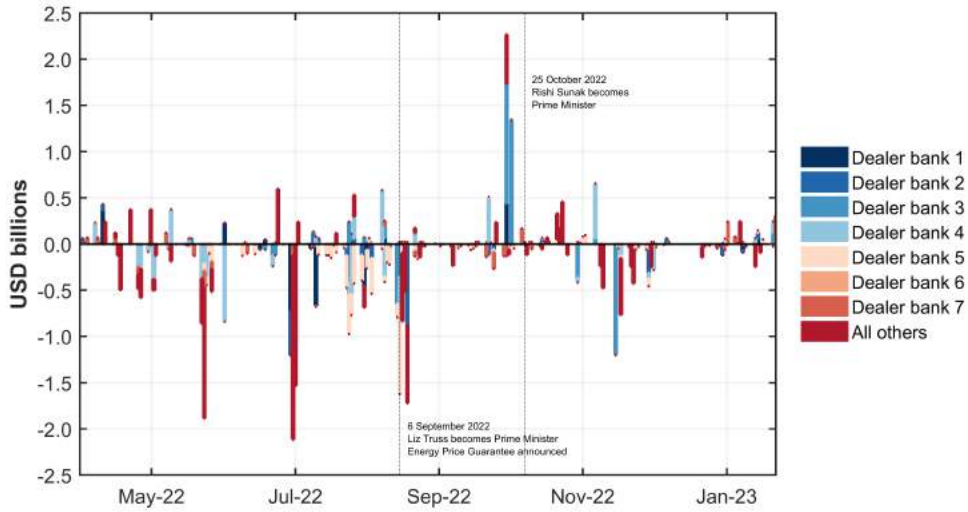
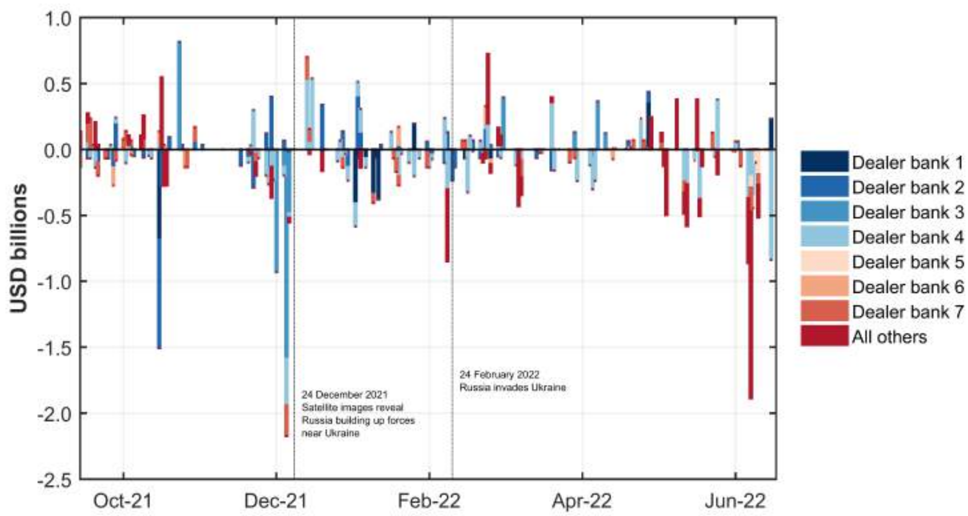


Figure E.20 Disagreement amongst dealer banks in the short horizon market

(a) From August 2022 (LDI crisis period)



(b) From February 2022 (Ukraine War period)



(c) From February 2020 (COVID-19 period)

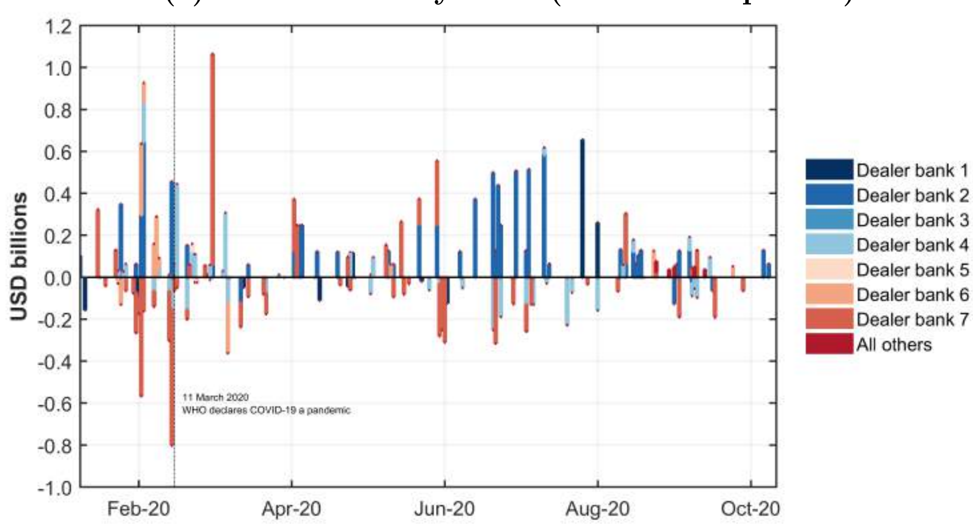
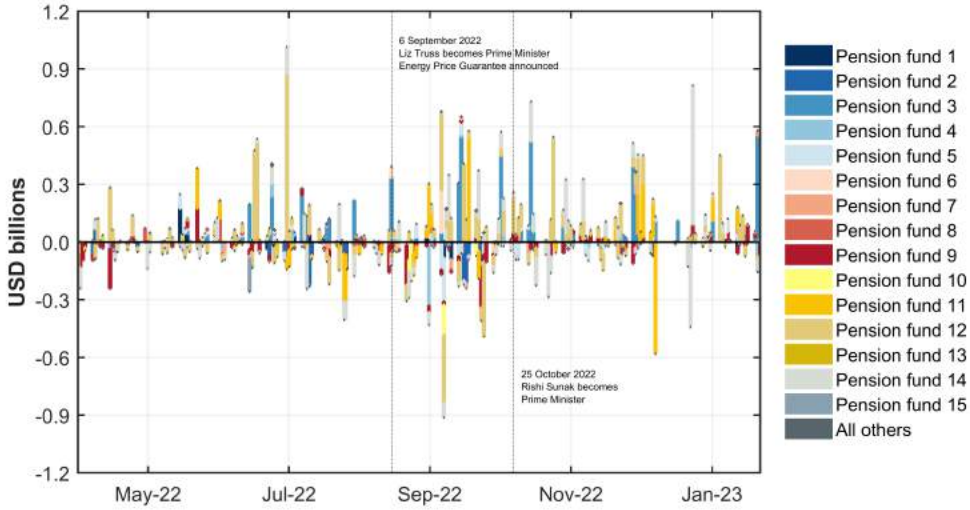
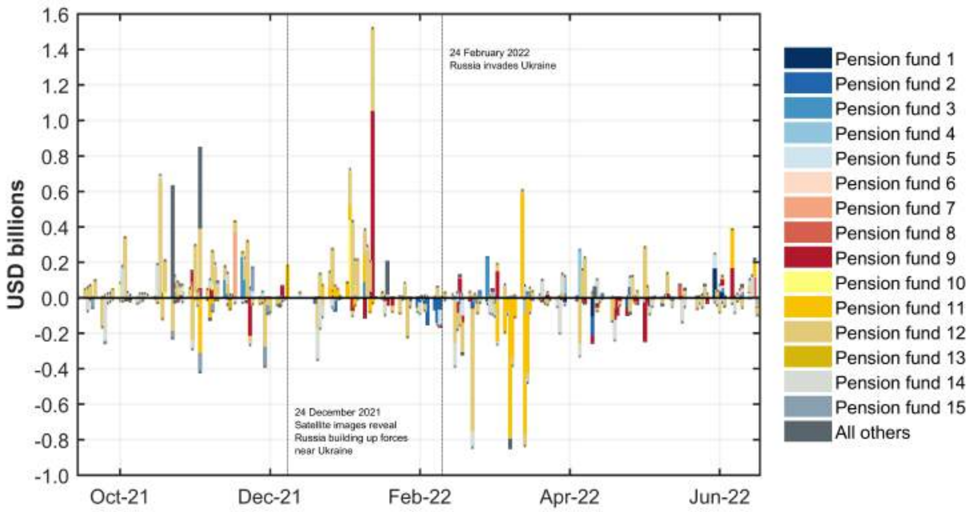


Figure E.21 Disagreement amongst pension funds in the long horizon market

(a) From August 2022 (LDI crisis period)



(b) From February 2022 (Ukraine War period)



(c) From February 2020 (COVID-19 period)

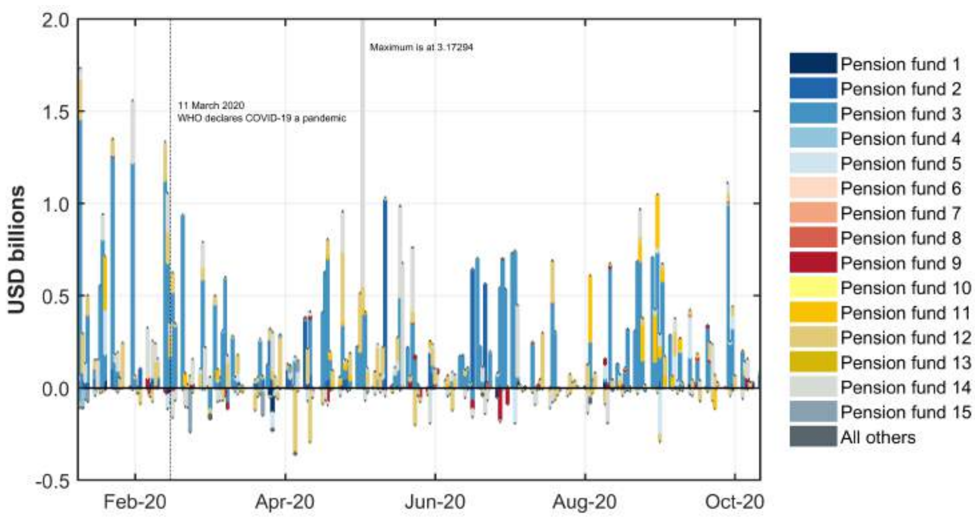
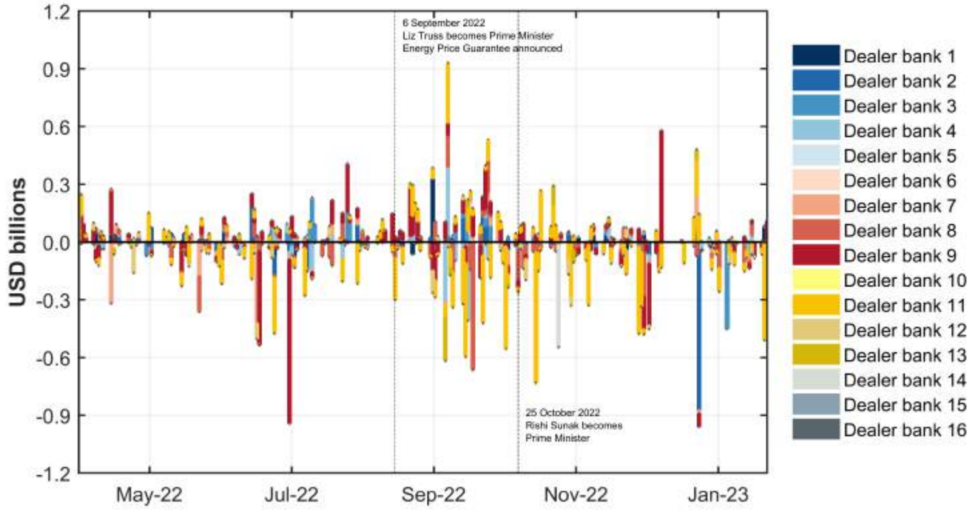
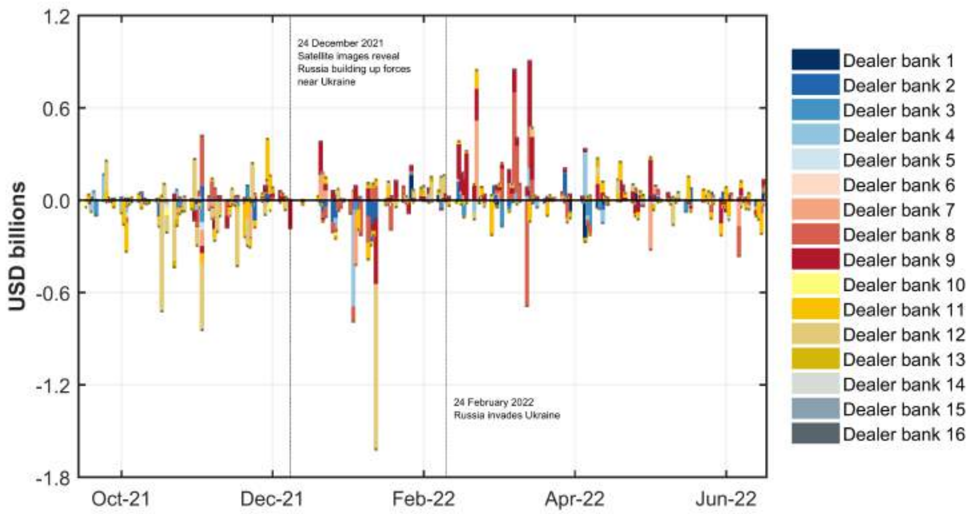


Figure E.22 Disagreement amongst dealer banks in the long horizon market

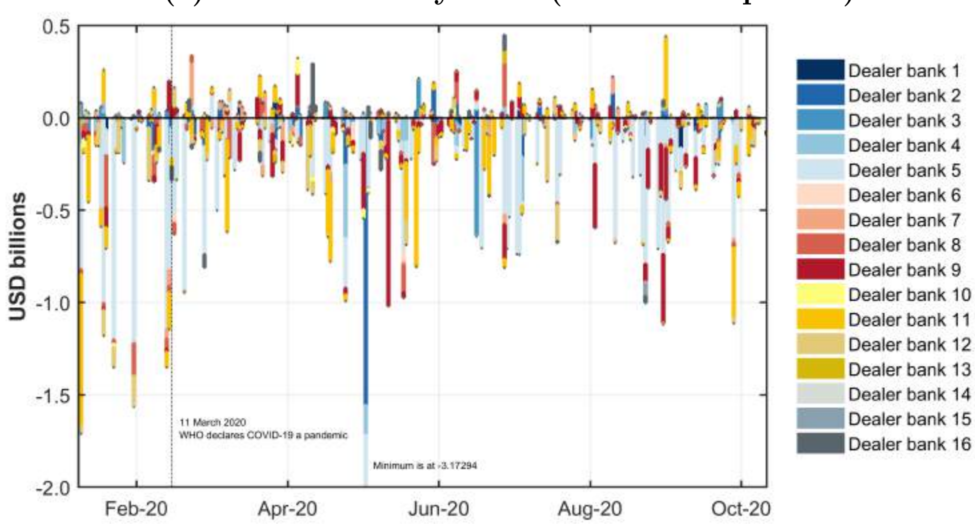
(a) From August 2022 (LDI crisis period)



(b) From February 2022 (Ukraine War period)



(c) From February 2020 (COVID-19 period)



F Robustness of estimation results

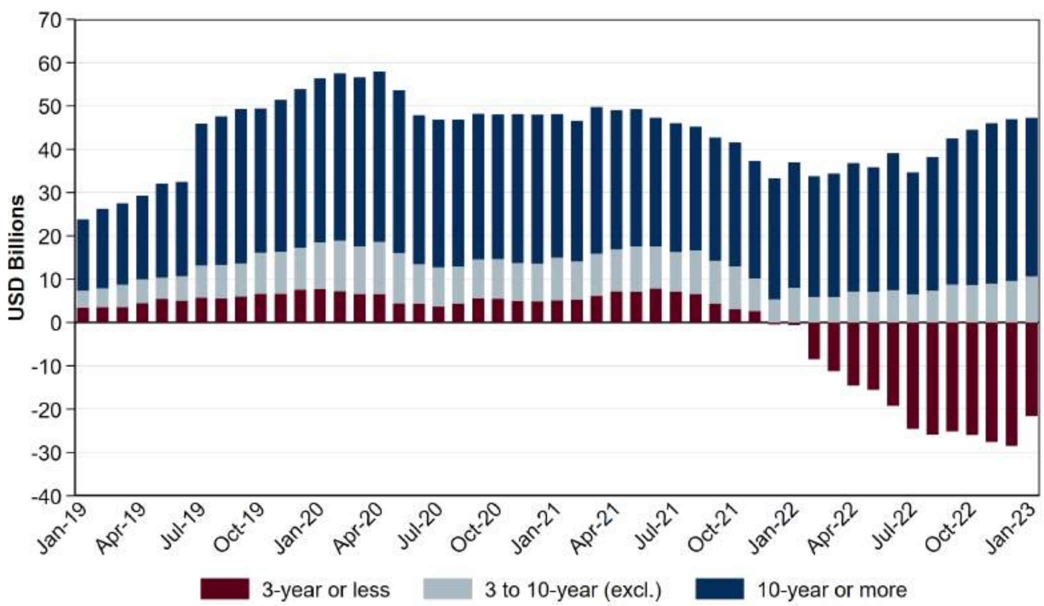
F.1 Adding insurance companies

In our baseline results, we focused on the trading volumes of hedge funds in the short-horizon market and PFLDIs in the long-horizon market. However, as discussed in Section 2 and also shown in Section B, another notable player 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.

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 F.1.

Figure F.1 Net notional position of insurers in UK RPI inflation swaps



Source: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.

This re-balancing is mechanical, and simply a 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, in the next sections, we show that our baseline results remain robust to the inclusion of insurers' quantities.

F.1.1 Estimated impulse response functions

In this section, we report the estimated impulse responses when insurance companies are added to the PFDLI sector.

The responses to the liquidity shocks under our first identification strategy — sign restrictions using the high-frequency data — are reported in Figure F.2. Comparing these estimates to Figure 12, 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 fundamental shocks are reported in Figure F.3 across the three identification strategies. Comparing them with the results in the main text given by Figure 11, again we find that the results are very similar whether insurance companies are included or not.

Finally, Figure F.4 reports the historical decompositions, in the form of the estimates time series for the three shocks driving liquidity premia in our model. Again the conclusions are similar to those in section 6.

F.2 Adding MPC announcement dates for heteroskedasticity-based identification

Figure F.5 Impulse response functions to a fundamental shock

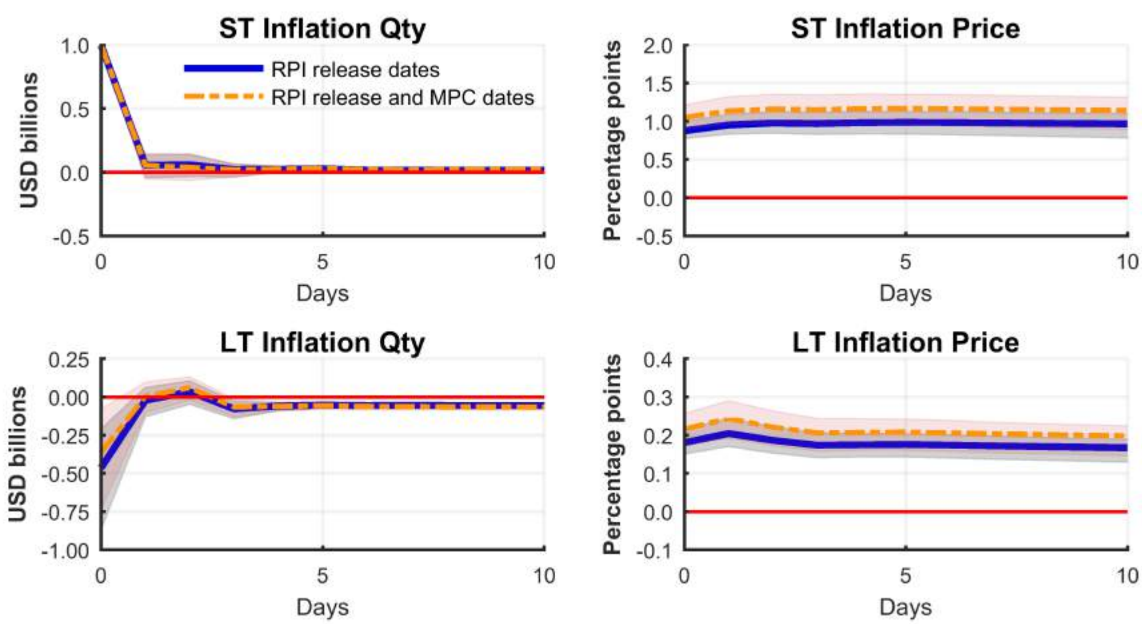
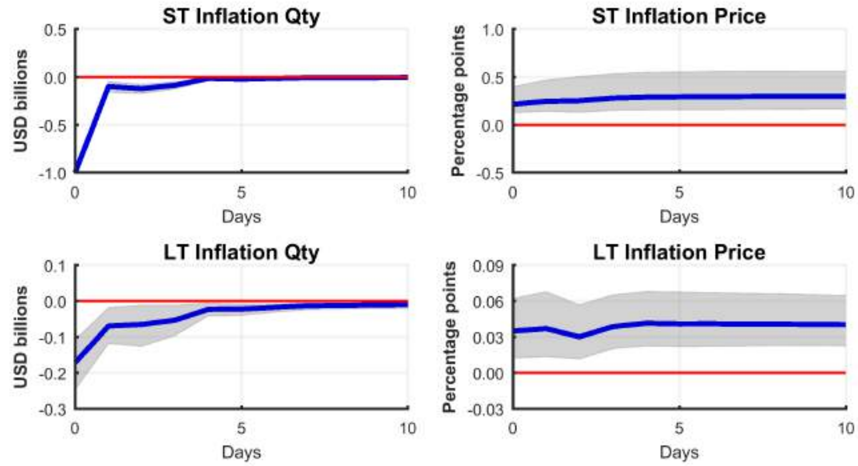
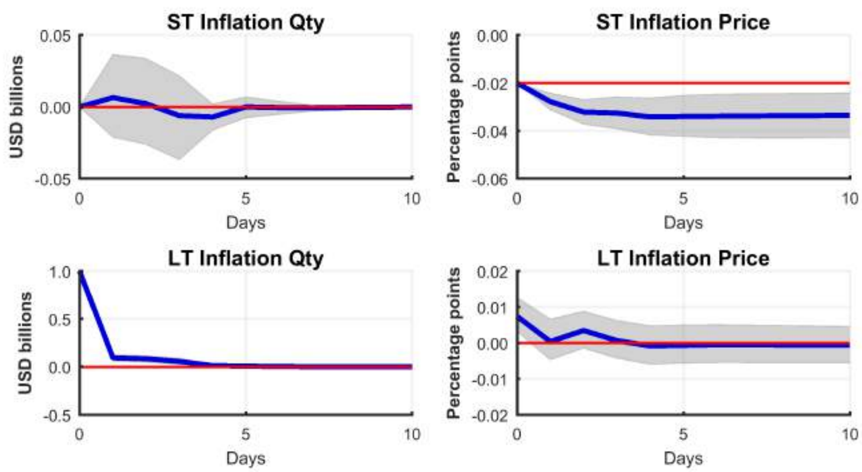


Figure F.2 Estimated impulse response functions to liquidity shocks

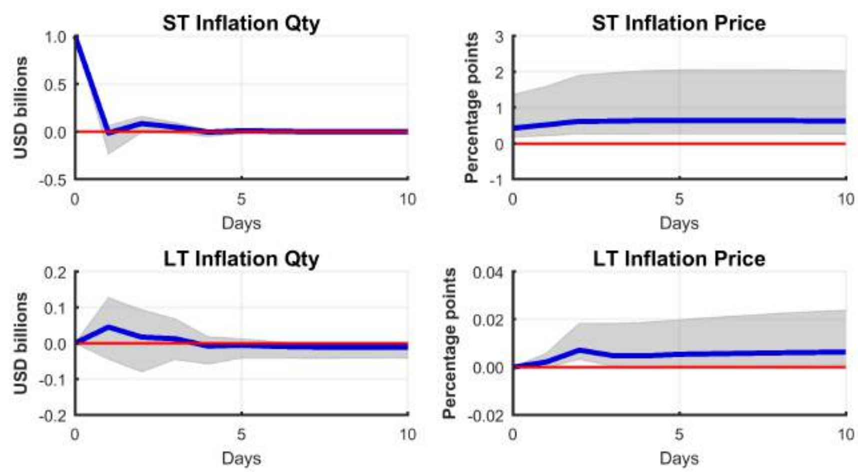
(a) Dealer supply shock ($\varepsilon_{b,t}$)



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



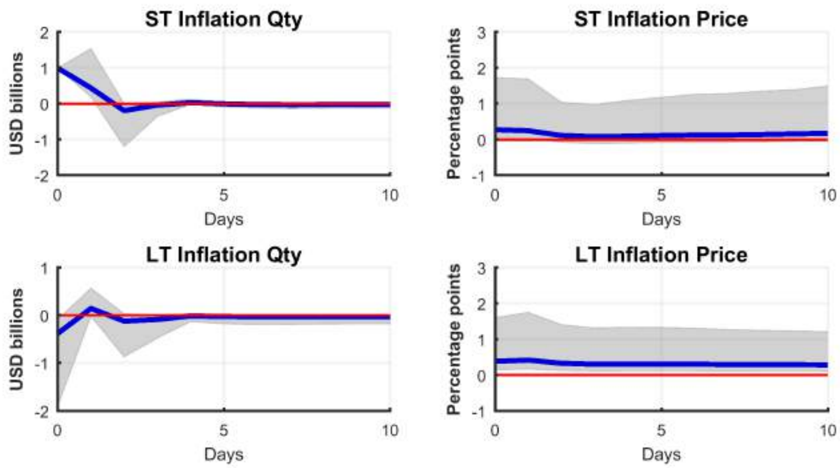
(c) Hedge fund demand shock ($\varepsilon_{h,t}$)



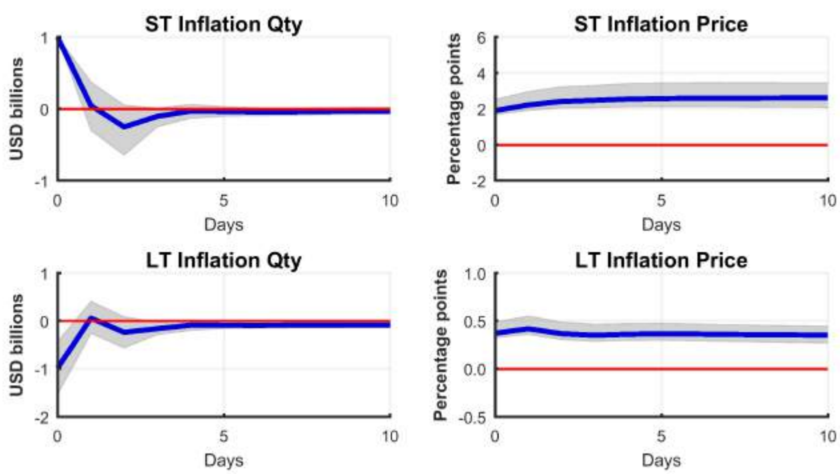
Note: Figure shows the median estimate with 68% confidence intervals.

Figure F.3 Estimated impulse response functions to a fundamental shock

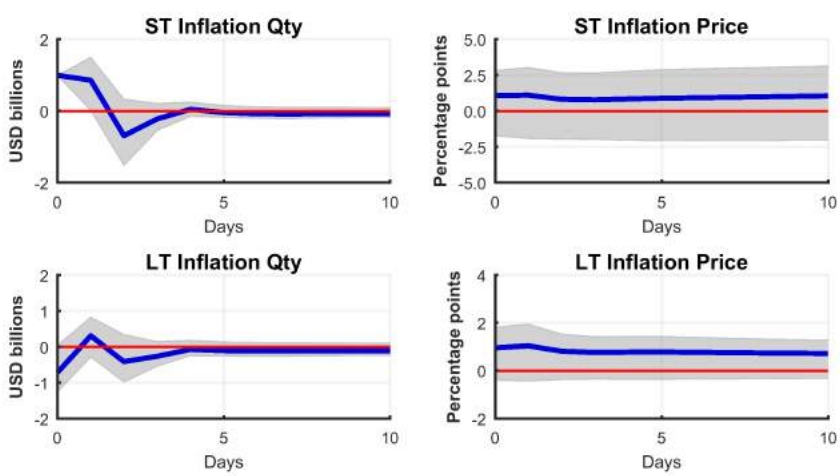
(a) With timing restriction strategy



(b) With heteroskedasticity strategy



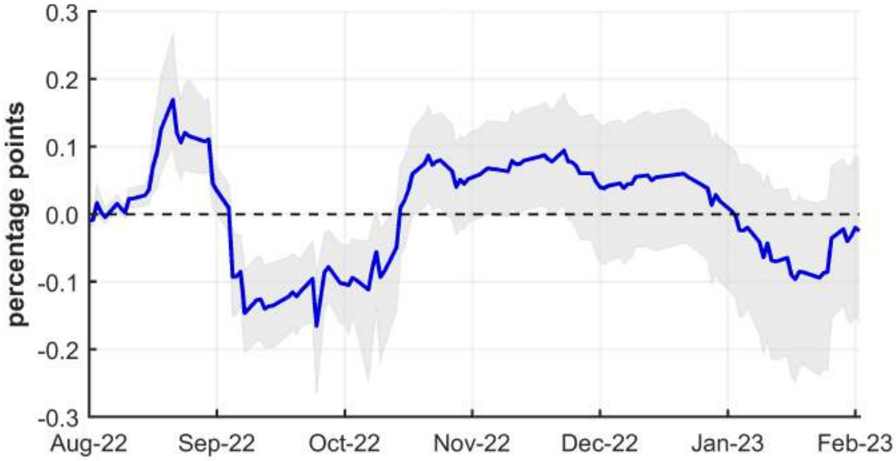
(c) With granular identification strategy



Note: Figure shows the median estimate with 68% confidence intervals.

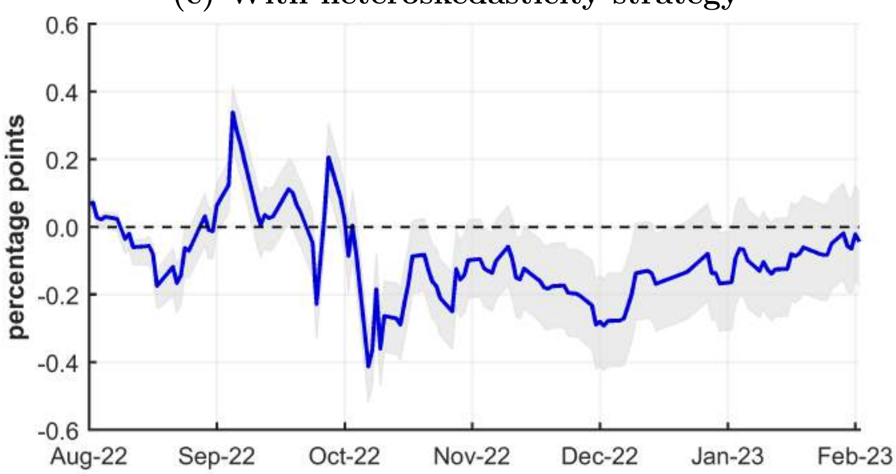
Figure F.4 Estimated liquidity shocks

(a) With timing restriction strategy



(b) With granular identification strategy

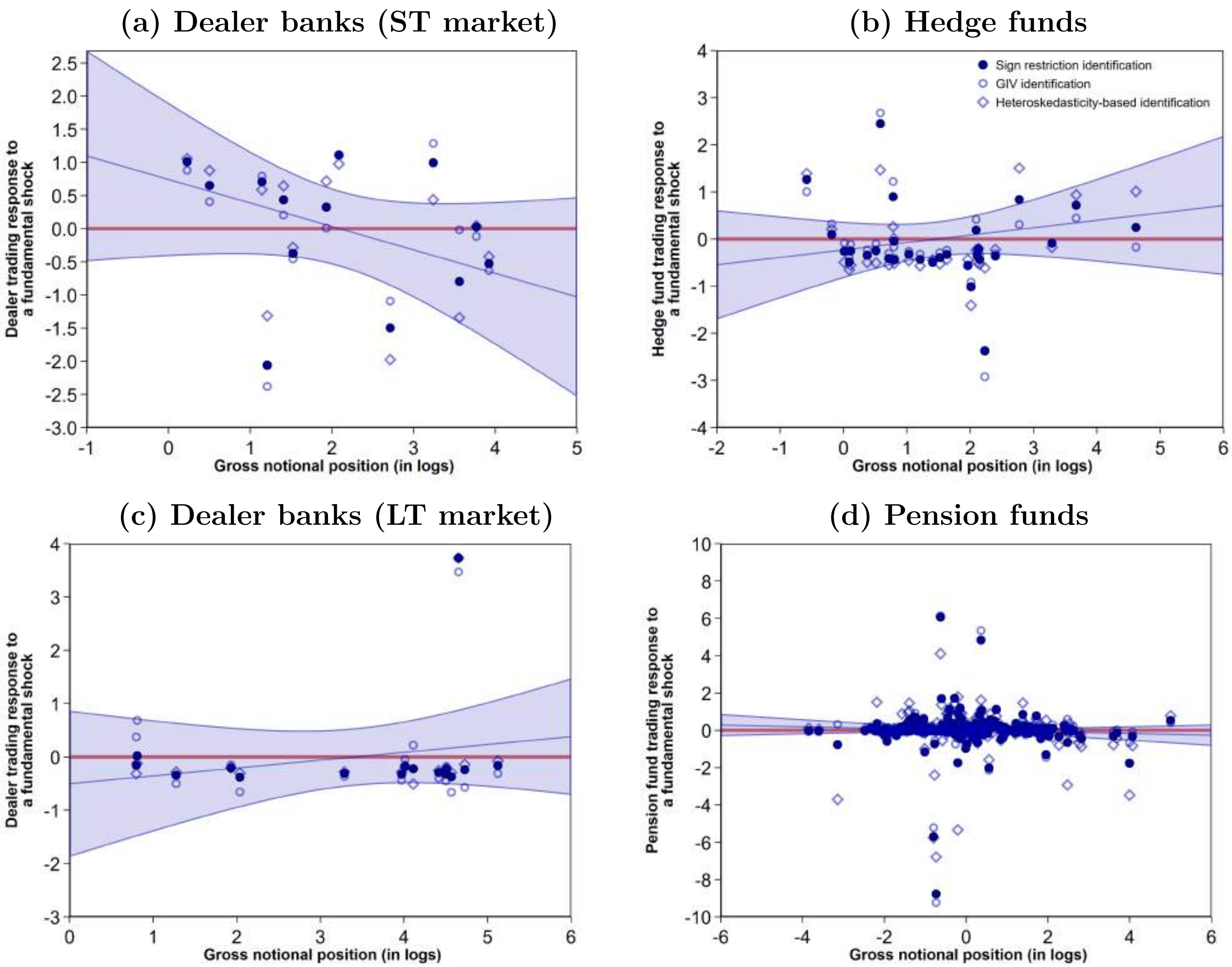
(c) With heteroskedasticity strategy



G Estimated responsiveness of trading position to fundamental shock

Figure G.1 show scatter plots of the estimates β_i against the size of the institution. Note the axis unit for pension funds is significantly smaller

Figure G.1 COEFFICIENTS SCATTERED AGAINST SIZE OF INSTITUTION



NOTE: The thick blue lines are pooled coefficient estimates with standard errors clustered at the institutional-level (where each is identified by a legal entity identifier), dashed blue lines are their 95% confidence intervals. SOURCE: DTCC Trade Repository OTC interest rate trade state files, from January 2019 to February 2023.