Bank of England

The market for inflation risk

Staff Working Paper No. 1,028
April 2025

Saleem Bahaj, Robert Czech, Sitong Ding and Ricardo Reis

This is an updated version of the Staff Working Paper originally published on 23 June 2023

Bank of England

Staff Working Paper No. 1,028

The market for inflation risk

Saleem Bahaj,[1] Robert Czech,[2] Sitong Ding[3] and Ricardo Reis[4]

Abstract

This paper uses transaction-level data on UK inflation swaps to characterise who buys and who sells insurance against higher inflation, and to introduce new measures of expected inflation. We find that this market is segmented: pension funds trade at long maturities, while hedge funds trade at short maturities, with dealer banks serving as the counterparty for both. We propose three novel identification strategies based on heteroskedasticity, granular instrumental variables, and sign restrictions to estimate demand and supply functions and to separate expectations from frictions in segmented financial markets. Expected inflation is more firmly anchored at long maturities than what swap prices indicate, while prices for short maturity swaps are mostly driven by shocks to frictions. Prices in this market absorb new information quickly, and the supply of long maturity inflation protection is very elastic. We find a strong correlation across institutions between their survey-based expectations and their trading behaviour.

Key words: Asset demand system, monetary policy, anchored expectations, identification of demand and supply.

JEL classification: C30, E31, E44, G12.

---

(1) University College London and Bank of England. Email: s.bahaj@ucl.ac.uk
(2) Bank of England. Email: robert.czech@bankofengland.co.uk
(3) London School of Economics. Email: s.ding5@lse.ac.uk
(4) London School of Economics. Email: r.a.reis@lse.ac.uk

The authors would like to thank seminar and conference audiences at the Bank of England, the Bank of Spain, Copenhagen Business School, City University Hong Kong, Danmarks Nationalbank, Duke University, Princeton University, University of Cambridge, University of Kent, University of Liverpool, the University of North Carolina Chapel Hill, University of Wisconsin, University of York, Hong Kong Baptist University, Fidelity, Fulcrum Asset Management, the NBER Summer Institute, the 3rd QCGBF Annual Conference, the 15th Annual Paul Woolley Centre Conference, the CESifo Venice Summer Institute, the 10th SAFE Asset Pricing Workshop, the 18th Conference on the Microstructure of Financial Markets, the 12th Bundesbank Term Structure Workshop, the 2024 CeMMAP Econometrics PhD Workshop, the Econometric Society 2023 European Winter Meeting, the 54th Annual Conference of the Money, Macro and Finance Society, the 2023 Sailing the Macro Workshop, the 25th Annual Macro Conference, the 11th HKEA Biennial Conference, the 2024 EFA meetings, the 2024

EEA-ESEM Congress, the 2024 ASSA meetings, the online Asian macrofinance seminar series, the 2023 CEPR symposium, as well as Matteo Aquilina, Nina Boyarchenko, Markus Brunnermeier, Jens Christensen, Valentin Haddad, Andrew Hauser, Kristy Jensen, Daniel Lewis, Daisuke Miyakawa, Aytek Malkhozov, Philippe Mueller, Carolin Pflueger, Fabricius Somogyi, Misa Tanaka, Dimitri Vayanos, Gyuri Venter, Motohiro Yogo and, in particular, Michele Piffer for their comments and suggestions. 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.

The Bank's working paper series can be found at www.bankofengland.co.uk/working-paper/staff-working-papers

Bank of England, Threadneedle Street, London, EC2R 8AH
Email: enquiries@bankofengland.co.uk

ISSN 1749-9135 (on-line)

1 Introduction

Investors wanting to trade inflation risk can do so by entering into swap contracts. Scores of papers in monetary economics have used the prices of these contracts as high-frequency measures of expected inflation and have studied how they respond to various shocks (e.g., Nakamura and Steinsson, 2018, Andrade and Ferroni, 2021). A vibrant literature in financial economics compares swap prices with those of inflation-indexed government bonds or with survey-based inflation expectations (e.g., Fleckenstein, Longstaff, and Lustig, 2014, Binder and Kamdar, 2022). Speeches by policymakers often discuss these market prices to anticipate the future path of inflation, and to assess whether expectations are anchored and monetary policy is credibly delivering price stability (e.g., Mann, 2022, Ramsden, 2022). Investors must form views on how swap prices correlate with both actual inflation and other asset returns in order to hedge against inflation and related macroeconomic risks (e.g., Cieslak and Pflueger, 2023, D'Amico and King, 2023).

In spite of their centrality to so many debates, there are reasons to be skeptical about swap prices as measures of expected inflation. After all, asset pricing research has found that prices often move for reasons unrelated to changes in the underlying payoffs. Inflation swap prices exhibit significant volatility, often fluctuating by tens of basis points within a short span of days, despite no discernible changes in economic conditions or monetary policy. This raises several research questions: how reliable are inflation swap price movements as indicators of expected inflation? How quickly do they incorporate information, and how do they compare with survey data? Behind these prices is a market where participants trade exposure to a central macroeconomic variable. Who are the market participants, and what motivates the transfer of inflation risk? What are the supply and demand curves for protection against inflation?

This paper brings new data and identification strategies to answer these questions. Using comprehensive transaction-level data on each over-the-counter (OTC) inflation swap contract traded within one of the world's largest inflation derivative markets, the UK, we identify the buyers and sellers of inflation insurance and to study how prices are formed. With a model of demand and supply, we decompose the price movements into a frictional component and a component capturing movements in risk-neutral expected inflation. Exploiting joint variation in both prices and quantities, we propose three distinct identification strategies for models of segmented financial markets that discern the shocks to these components through time and across different maturities. We then use the data on quantities to clean the price data, thereby producing a new measure of expected inflation. To understand how inflation risk is transferred and traded, we estimate the demand and supply for inflation protection, the relative price impact of different investors, how frictions relate to market liquidity, and how institutions' surveyed expectations of inflation correlate with their trading behavior.

First contribution: facts and market structure. We provide a description of a large and important financial market that is opaque because of its OTC nature. This enables us to identify the entities influencing inflation swap prices, understand their motivations for trading inflation risk, and determine the appropriate model to interpret this market.

Section 2 establishes three stylized facts on the UK market for inflation swaps. First, dealer banks are not neutral market makers. Rather, they are large net sellers of inflation protection. Their net positions significantly exceed their holdings of index-linked bonds, the closest hedge. Moreover, at the institution level, there is little correlation between trading activity in the index-linked bond and inflation swap market. Banks are more accurately characterized as sellers of inflation insurance rather than merely as market makers with a matched book.

Second, at long maturities (10 years or more) the buyers of inflation insurance are mostly pension funds. They hold large, persistent, positive net positions, and most of the variation in trading volume in that maturity segment is driven by their actions. Consequently, when inflation rises unexpectedly, there is a direct flow of payments from dealer banks to pension funds.

Third, at short maturities (3 years or less), hedge funds, acting as informed traders, hold small net positions against dealer banks. However, they actively trade, and their net position fluctuates between positive and negative on any given day. On average, they sold inflation protection when inflation fell during the pandemic and bought protection ahead of the recent inflation spike.

In other words, we observe a remarkable *segmentation* in this market: pension funds maintain large long maturity swap positions while barely trading in the short maturity segment, while informed traders conduct most of their trading activity at short maturities. Dealer banks are active in both. Section 3 proposes a model of this market split in two different segments based on the maturity of the swap. Neither demand nor supply curves are horizontal because all parties are averse to risk, have limited funds, and hold other positions and income that correlate with the realizations of inflation. The outstanding quantity of inflation protection is non-zero because institutions have different risks to hedge, different beliefs about future inflation, and different capacities to bear inflation risk.

Second contribution: identification strategies. If all participants uniformly expect higher inflation, then both supply and demand curves in the market will shift vertically by the same amount. The price will rise one-to-one with expected inflation. If institutions instead update their beliefs differentially, then the price increase reflects a risk-aversion and size-weighted average of the change in expected inflation. Likewise, if all participants revise upward their beliefs regarding the volatility of inflation or its covariance with market returns, both supply and demand shift upwards, and the price increase reflects the rise in the inflation risk premium that policymakers and researchers care about, and so in risk-neutral expected inflation.

However, consider a scenario where pension funds experience a shift in the composition of their liabilities between defined benefit and defined contribution elements, or there is a change in the shadow cost of posting collateral for a swap. Then, the demand curve for long maturity inflation swaps will shift, and so will the price, despite no change in expected inflation. Similarly, if dealer banks experience changes in their balance sheet capacity to supply inflation protection due to losses or gains in other business areas, or if there are regulatory changes regarding the capital they must hold against their positions, then the supply curve alone will shift, moving the price as well.

Deriving measures of expected inflation from swap prices requires filtering out these friction-driven changes in how one side of the market trades and prices risk. This creates a difficult identification problem of separating shocks to fundamentals from shocks to frictions or, equivalently, of estimating the supply and demand functions. Section 4 contributes by showing how three identification strategies can be applied to segmented financial markets to accomplish this task, each exploiting distinct sources of variation in the data.

The first strategy exploits the time-varying volatility in the time series. Regular releases of official inflation statistics cause heightened volatility in the inflation swap market. Assuming that shocks to expected inflation drive this volatility, the heteroskedasticity identifies these shocks.

The second strategy instead exploits the cross-institutional variation in trading. The transaction-level data is granular: the size of institutions' positions follows a power law distribution. We estimate institution-level disturbances and use them to build three instrumental variables for the movements in supply and demand that correspond to frictional shocks.

The third strategy exploits the high frequency of the data. The identifying assumptions are twofold: first, the desks responsible for trading short maturity and long maturity swaps within a given dealer bank operate independently over the course of a single day; and second, within the same time frame, hedge funds respond more to fundamental factors compared to dealer banks, which in turn are more responsive than pension funds. These assumptions impose restrictions on the relative shifts of demand and supply functions in response to shocks that amount to zero and sign restrictions on the structural responses of prices and quantities to the shocks.

Third contribution: empirical estimates. Section 5 reports estimates of UK expected inflation between January 2019 and February 2023 that are consistent across the three identification strategies.

How reliable are swap prices in comparison? Our first finding is that, at long maturities, the observed swap price tracks expected inflation reasonably well; however, discrepancies between the two series occasionally were as high as 30 basis points. Second, inflation swap prices overstated movements in expected inflation in several episodes. For example, they overstated the risk of deflation during the pandemic, while they overstated the inflationary pressures during the energy crisis triggered by the invasion of Ukraine. Furthermore, from September 2022 to February 2023, the prices overstated the degree to which inflation expectations were unanchored. Third, at short maturities, the frictions dominate, and swap prices can diverge substantially and persistently from expected inflation.

Section 6 turns to how this market incorporates information and shares inflation risk among its participants. We find that the impulse responses become horizontal within one to three days. The inflation swap market, therefore, seems to incorporate new information relatively quickly. Moreover, the supply of inflation protection by dealer banks to pension funds at long maturities is very elastic, unlike their supply to hedge funds at short maturities. This indicates that banks play a crucial role in shouldering changes in demand for long maturity inflation insurance. Finally, we find that most of the variation in the price of short maturity inflation swaps is driven by frictional movements. Supply shocks from dealer banks explain around a third of the variation in short maturity inflation swap prices; the majority is instead due to demand shocks from pension and hedge funds, which have been less explored in the existing literature.

Section 7 matches the positions of financial institutions in the inflation swap market with their answers to surveys on expected inflation. We find a significant positive correlation between the two. This is a rare validation of a link between survey-based measures of expected inflation and actual trading behavior, with large sums at stake.

Connections to the literature. Beyond the vast literatures on inflation, expected inflation, monetary policy, and financial markets that we already discussed, this paper directly builds on a few other papers.

First, we highlight a particular market where the segmentation across maturities is stark, building on the work of Vayanos and Vila (2021). While this market for inflation insurance is not as large as the markets for bonds or foreign exchange that the prior literature has focused on, it is very significant for macroeconomic outcomes and monetary policy. With transaction-level data on a large sample of the market, we can empirically identify the arbitrageurs and preferred-habitat agents described in these 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 et al. (2025). While previous studies have used data on stocks (Koijen et al., 2024), bonds (Koijen et al., 2021) or exchange rates (Koijen and Yogo, 2020), we focus on the large and liquid OTC 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, 2024), we build on Gabaix and Koijen (2021) but use trade-level positions across different investor types to build an instrument for the demand of each type of investor.

Third, a literature following Begenau et al. (2015) has looked at the interest rate swap market to identify the sectors bearing interest rate risk and how their exposure varies. We study inflation derivatives, and we directly observe the directional positions taken on by dealer banks and other investors. In contrast to the findings of Granja et al. (2024) and McPhail et al. (2023) on interest rate swaps, we find that banks take large net positions in the inflation risk market. We also find that demand shocks affecting the clients of the dealers play an important role in driving market prices, alongside shocks to dealers themselves. Like Hanson et al. (2024), 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 about the drivers of inflation swap prices.

Fourth, numerous studies, including recent contributions by Fang et al. (2025) and Boons et al. (2020), utilize market price data to separately estimate subjective expected inflation and inflation risk premia. Our study, however, focuses on disentangling risk-neutral expected inflation—which encompasses both subjective expectations and risk premia—from market frictions such as liquidity and other premia. Conversely, Reis (2020) pools market prices and survey data to derive measures of expected inflation. In contrast, we employ micro-level data on trading behavior and survey responses to check for consistency between the two.

Fifth, and finally, we use the regulatory EMIR Trade Repository (TR) data on trade-level OTC derivative positions (for an overview, see e.g. Abad et al., 2016). Cenedese et al. (2021) and Hau et al. (2021) use the EMIR TR data on FX forwards and swaps to investigate the impact of the leverage ratio and price discrimination, while Cenedese et al. (2020) use data on the interest rate swap market to compare prices in OTC transactions with those in centrally-cleared trades.[1] To the best of our knowledge, our work is the first to use the granular EMIR TR data on inflation swaps.

2 Data, summary statistics, and stylized facts

We begin by describing an inflation swap, followed by an explanation of our data sources. Finally, we establish three key facts that characterize this market as segmented into two parts.

2.1 Inflation swap contracts

An inflation swap is a bilateral contract in which one party, the floating leg payer, pays the realized cumulative growth in a price index over the contract's duration. In return, the other party, the fixed leg payer, pays a fixed amount at a pre-agreed rate. This fixed rate is known as the breakeven inflation rate because, if the realized inflation rate matches the fixed rate at the contract's conclusion, both counterparties break even. This fixed rate is the *price* of the swap. It ensures that the net present value of the swap for both counterparties is zero at initiation. Consequently, no cash flows are usually exchanged at the start of the contract. As the floating leg payer is selling protection against high inflation in exchange for a premium that is the fixed rate, this makes swaps an attractive way to buy and sell inflation insurance.

Most swaps are zero-coupon, meaning that payments on the fixed and floating legs are exchanged only at the end of the contract. However, the two parties continuously exchange payments, known as variation margin, to ensure that the contract remains at zero net present value.

[1] Concurrent with our paper, but focusing on the FX derivative market, see Czech et al. (2022) and Ferrara et al. (2022). Additionally, building on our work, see also Khetan et al. (2023) and Jansen et al. (2024).

Therefore, the *payoff* of an ongoing swap is the exchange of variation margin. Since this payoff corresponds to the net change in the value of the swap position, it is nearly equivalent to the position being closed at day-end, with the agent whose side of the contract appreciated receiving a payment. Consequently, payments are tied to changes in expected inflation, even if it will take years for the actual underlying inflation to be realized. This also means that there is minimal counterparty risk because each party is nearly indifferent between maintaining the swap contract or closing it and keeping the accrued variation margin.

Each swap contract specifies an amount swapped that links the size of cash flows to inflation and the fixed rate, referred to as the gross notional. To give an example, imagine a \$100 gross notional swapped for 1 year with a fixed rate of 1%. If inflation was 1% over the year so that prices rose by 1%, both counterparties would break even. If there was no inflation over the course of the year the fixed rate payer would pay \$1 to the floating rate payer, but if inflation was 2% instead the floating rate payer would pay \$1 to the fixed rate payer. We use net notionals, which consolidate all the exposures of an institution, to measure the *quantity* of inflation protection held by that institution. This measure is positive if the institution is a net buyer and negative if it is a net seller. For example, if an institution holds a gross \$100 of swaps as the floating leg payer and \$200 as the fixed leg payer, its net notional position is \$100.[2]

2.2 Inflation-linked debt and inflation swaps

Another method to obtain inflation protection is through holding inflation-linked government debt, commonly referred to as linkers in the UK. However, linkers are less attractive for hedging against or speculating on inflation due to upfront capital costs and illiquidity during periods of market stress. Additionally, the supply is limited, with bonds outstanding amounting to only approximately 20% of GDP during 2019-2022. In contrast, the total gross notional amount of UK inflation swaps outstanding during the same period ranged between 110-130% of UK GDP. As demonstrated in Appendix A.1, there is minimal correlation at the institutional level between trading in linkers and swaps.

For historical reasons, the retail price index (RPI) is typically used to calculate the inflation coupon on linkers. Since the linker market predates the inflation swap market, the latter has also adopted the convention of using RPI to compute cash flows on the floating leg. RPI inflation has typically been approximately 1.5 percentage points higher than the consumer price index (CPI) inflation, which is the basis for the Bank of England's 2% target.[3]

[2] There are some complexities regarding the timing of the contract that we discuss in a supplemental Appendix.

[3] The supplemental Appendix further elaborates on this discrepancy and compares UK price indices.

2.3 The EMIR trade repository data

Inflation swaps are traded in an OTC market, where the terms of a transaction are negotiated between the two counterparties rather than on an exchange. By regulation, UK legal entities have to report the terms of any derivative transaction to an authorized trade repository by the next business day. Initially mandated under the European Market Infrastructure Regulation (EMIR), reporting requirements were later adopted in UK legislation under UK-EMIR. Our data consists of all trades submitted to DTCC Derivatives Repository Plc ('DTCC') in which at least one of the counterparties is a UK-regulated entity. This is the largest trade repository in terms of market share and is representative of the overall market.

We use the DTCC's daily reports to capture the stock of all outstanding inflation swap contracts on a given day, as well as to obtain the flow of trading activities. Since the vast majority of trades are spot contracts (rather than forward contracts), we focus on these. We clean the data by eliminating: (i) duplicated transaction-level reports, (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. We also drop observations with inconsistent values for the reported notional, the identities of the counterparties, the counterparty side, the maturity date, or the underlying inflation index. This leaves us with more than 25 million unique trades, from the 2^{nd} of January 2019 to the 10^{th} of February 2023. We allocate investors to groups using a best-endeavor sectoral classification.[4]

2.4 The dealer-client segment of the market

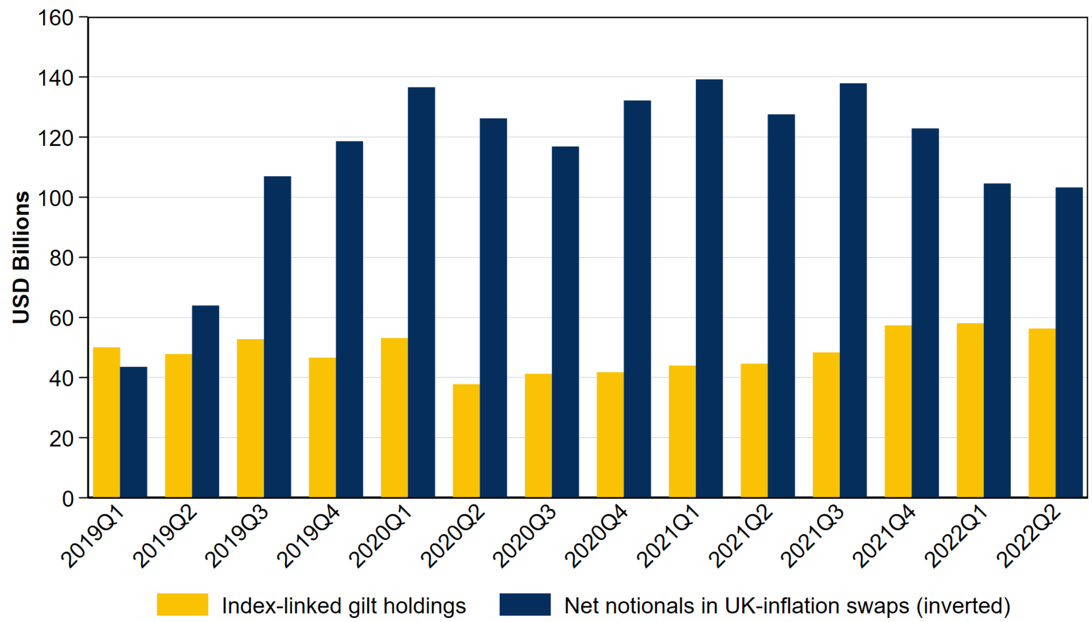
The total gross notional outstanding in the UK inflation swap market fluctuated between \$3.5tn and \$4tn during 2019-2022. Our focus is on how these swaps are utilized to transfer inflation risk across different segments of the financial system. However, approximately 60% of trades involve a central clearing counterparty (CCP). Only clearing house members—predominantly large international banks that supply swap contracts to the broader financial system and that we refer to as dealers—can trade directly with a CCP. Consequently, trades with clearing houses do not provide insights into the types of institutions that demand and trade inflation risk.[5] Additionally, a further 17% of trades occur in the intra-dealer market, only involving transactions between dealers.

We focus on the remaining 22% of contracts, equivalent to \$1.1tn in gross notional terms, which are sold by the 16 dealers to their clients. A third of dealer-client trades by value are with pension funds and liability-driven investment funds (collectively referred to as pension funds). A third are with well-informed active traders (hereafter, hedge funds), and the remaining third are with other

[4]The supplemental Appendix details the trade repository data and the procedure to clean the data. It also compares our dataset with supervisory data on derivatives holdings within the insurance sector, revealing a close match.

[5]A client trading with a dealer can elect to have that trade cleared through a CCP. In our data, this would manifest as two transactions: one between the dealer bank and the CCP and another between the dealer and the client. Our interest is in the latter transaction, and by focusing on the dealer-client market, we avoid double counting.

Figure 1 FACT 1: DEALERS HAVE A NET EXPOSURE TO HIGHER INFLATION



NOTE: The figure shows in blue the aggregate net notional position of dealer banks in the dealer-client market segment of the UK inflation swap market across all maturities at the end of the quarter. A positive value indicates that the dealers are net *sellers* of inflation protection. In yellow-gold are, for the same dealer banks, the aggregated face value of holdings of index-linked government debt. SOURCE: UK bank granular exposures data (Covi et al., 2022) and DTCC Trade Repository, from March 2019 to June 2022.

non-banks, such as insurance companies, non-financials, asset managers, and sovereigns (which we exclude). Within the dealer-client segment, approximately two-thirds of swap contracts are traded at long maturities of ten years or more, although there is a significant market for shorter maturities of one to three years.[6]

2.5 Three facts on the UK inflation swap market

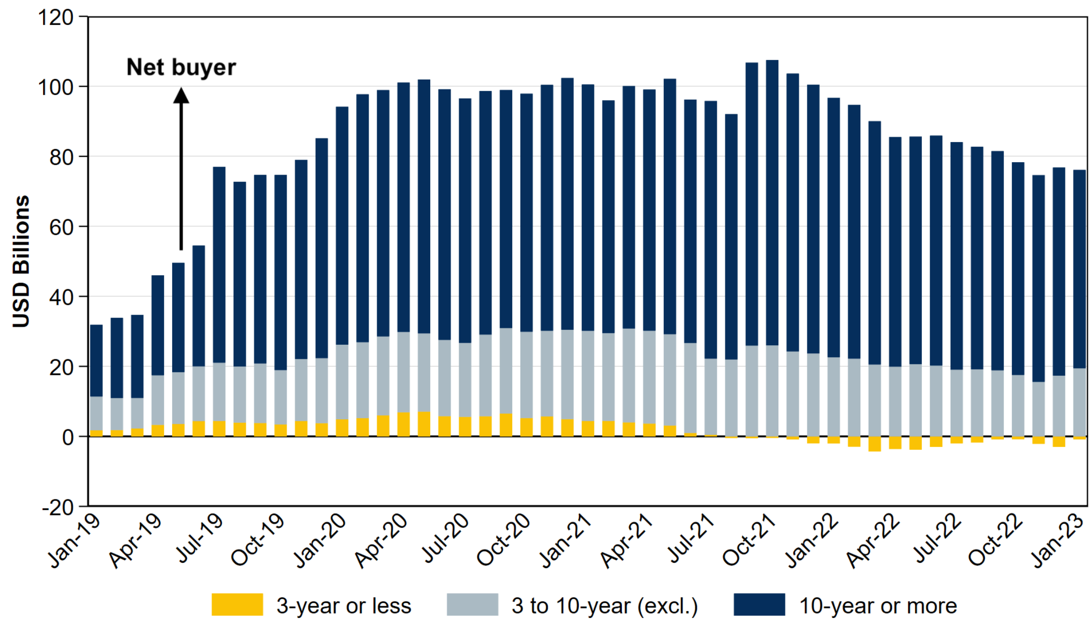
Fact 1: Dealers are not neutral market makers. Figure 1 shows that dealer banks have sold inflation protection in the swap market beyond their holdings of linkers, the direct hedge against inflation risk, since at least the second half of 2019.[7] The magnitude of their net exposure to inflation, via linkers and swaps, reached a peak of almost \$100bn in 2021. This implies that the 15% cumulative overshoot in UK inflation over 2021-2023 may have cost UK banks around \$15bn, or about 3% of their capital, in pure cash flow terms. This phenomenon is consistent across the entire sector: nearly all dealer banks sell inflation protection and thereby take on inflation risk.

Fact 2: Pension funds buy protection from dealers at long maturities. Pension funds have substantial real liabilities in the form of index-linked pension commitments to their members, and

[6]The supplemental Appendix provides further information on gross notionals by maturity and institution, over time, and at a single date. It also describes the role of UK insurers and explains why we categorize them as others.

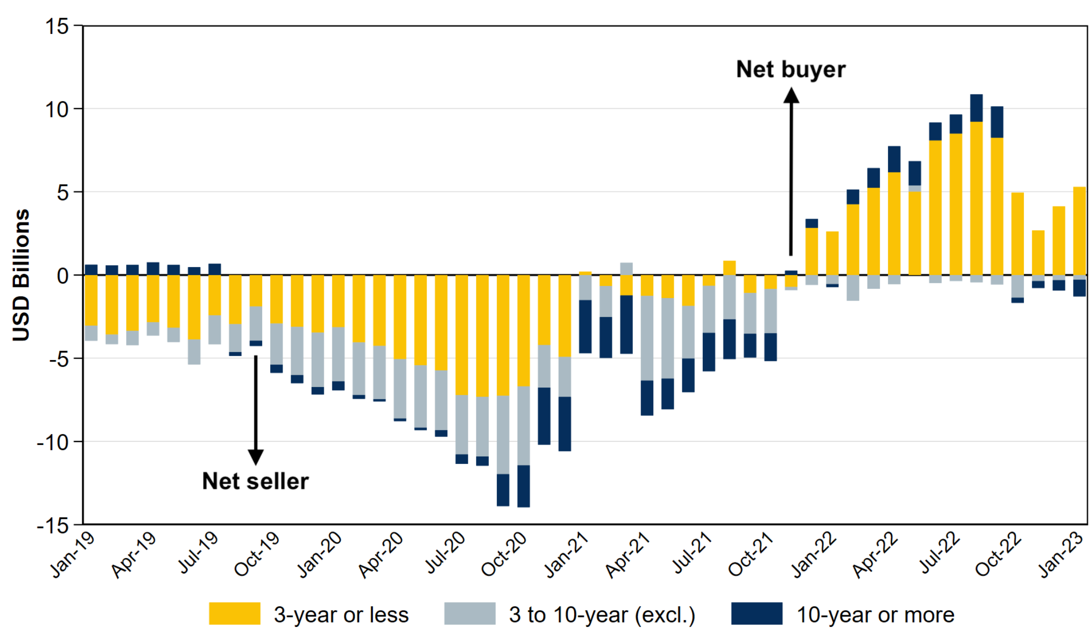
[7]This difference is not explained by differences in maturities, which are quite similar for the dealer banks' holdings of linkers and swaps. Of course, dealers will manage their inflation exposures through other correlated asset classes, like commodities, and our model will explicitly consider their portfolio strategies.

Figure 2 Fact 2: Pension funds buy long maturity protection from dealers



Note: The figure shows the aggregate net notional position of pension funds *vis-à-vis* dealers in the UK inflation swap market at month-end, categorized by the maturity of the underlying swap contract. A positive value indicates that the funds are net buyers of inflation protection. The data sample is from January 2019 to February 2023. Source: DTCC Trade Repository.

Figure 3 Fact 3: Hedge funds are active in the short maturity market segment



Note: The figure shows the aggregate net notional position of hedge funds *vis-à-vis* dealers in the UK inflation swap market at month-end, categorized by the maturity of the underlying swap contract. A positive value indicates that the funds are net buyers of inflation protection. The data sample is from January 2019 to February 2023. Source: DTCC Trade Repository.

the limited supply of linkers covers only a small portion of these liabilities.[8] It is not surprising

[8]The UK has a large defined benefit pension fund sector. The hedging demand from these pension funds is the primary reason for the deep market of inflation-linked bonds and derivatives.

that, in our data, they are consistently large buyers of inflation protection from dealers, with a maximum position reaching approximately $110bn in our sample.

Figure 2 breaks down their net notional positions by the initial time-to-maturity of the contracts. Strikingly, they almost exclusively hold inflation swap contracts with an initial maturity of 10 years or longer, likely reflecting the long duration of their liabilities.

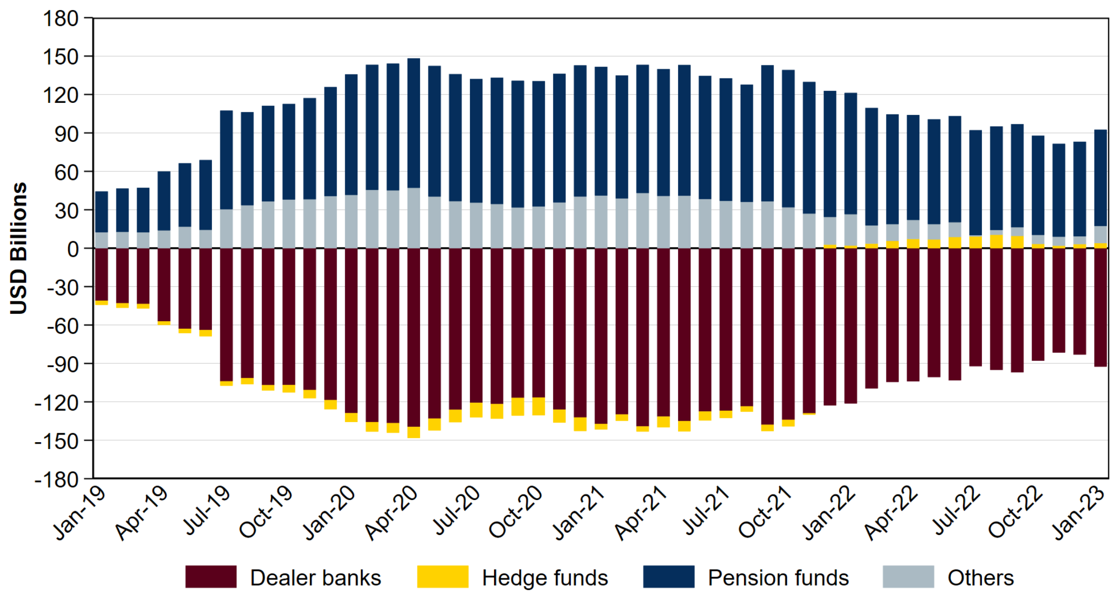
Fact 3: Hedge funds trade inflation risk primarily in the short maturity market. Figure 3 shows the net notional positions of hedge funds. These positions are substantially smaller than those of pension funds, reflecting the absence of an implicit desire for insurance, unlike pension funds. However, these small net positions conceal larger gross positions, as detailed in Appendix A.2, and significant divergence within the sector, with different institutions taking opposing sides at any given time, indicative of speculative behavior regarding inflation levels. On an aggregate basis, the sector's net positions are small because different institutions are often on opposite sides of the market. Overall, the sector transitioned from being net sellers to net buyers of protection as inflation initially declined and then increased. These institutions predominantly trade inflation swaps with an initial maturity of 3 years or less.

2.6 Market segmentation

Figure 4 presents the complete picture for the net notionals across different investor types. Combined, the three facts imply a remarkable *segmentation* of the UK inflation swap market: pension funds primarily trade in the long maturity market segment, where they hold persistently large positive net positions and hence buy inflation protection; hedge funds trade in the short maturity market segment with small and fluctuating net positions (but large gross positions and significant disagreement among individual funds); and dealer banks are the counterparties in both market segments to both types of clients, trading actively in both and overall acting as sellers of protection. From the perspective of segmented markets models (e.g., Vayanos and Vila, 2021), dealers serve as arbitrageurs across maturities, while pension funds and hedge funds are preferred habitat investors in the long maturity and short maturity market segments, respectively. Appendix A.3 finds the same pattern in the Euro-inflation swap market, albeit with a shorter sample period.

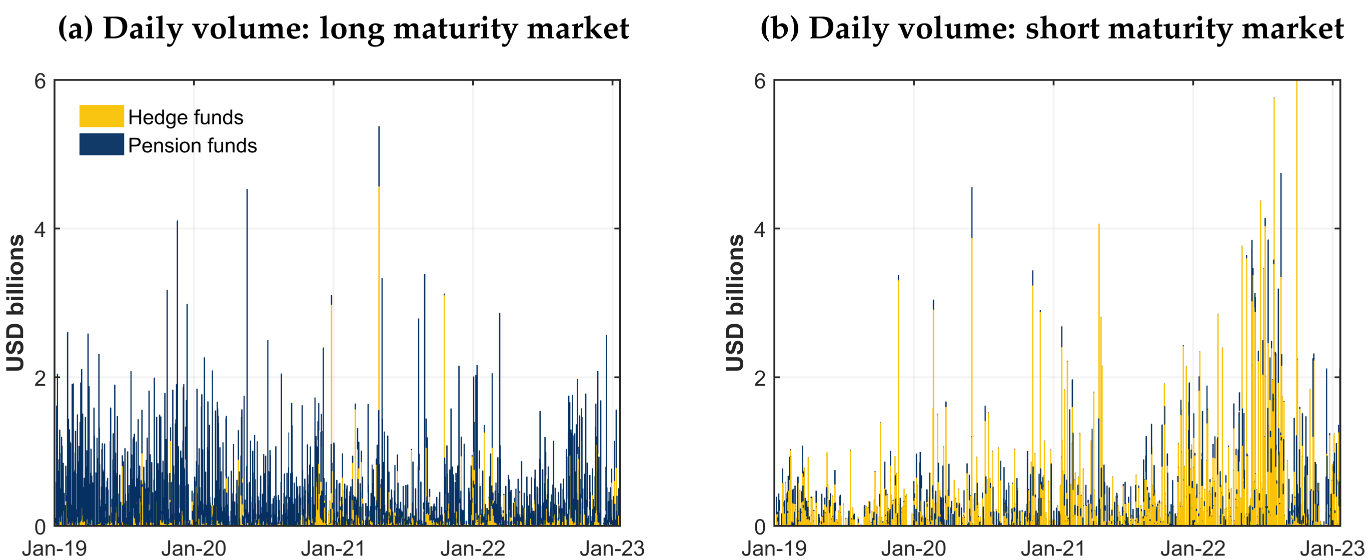
Transitioning from net positions to high-frequency trading behavior, Figure 5 illustrates the gross notional value of new contracts initiated by pension funds and hedge funds, aggregated daily across participants and categorized by maturity. Trading volumes in both markets exhibit similar magnitudes. However, pension funds' gross trading volume in the long maturity segment significantly surpasses that of hedge funds, whereas hedge funds dominate the short maturity segment. The median volume-weighted maturity of swap contracts executed by pension funds is 11 years, compared to 2 years for hedge funds, as detailed in Appendix A.4. Although there is some trading activity by hedge funds in the long maturity segment and by pension funds in the short maturity segment, these activities are relatively minor compared to the trading of the

Figure 4 NET NOTIONAL POSITION BY INVESTOR TYPE IN THE INFLATION SWAP MARKET



NOTE: The figure shows the aggregate net notional position of all sectors in the dealer-client segment of the UK inflation swap market at month-end. A positive value indicates that the sector is a net buyer of inflation protection. The data sample is from January 2019 to February 2023. SOURCE: DTCC Trade Repository.

Figure 5 MARKET SEGMENTATION IN TRADING AT HIGH FREQUENCIES



NOTE: The panels illustrate the gross notional value of new contracts initiated by hedge funds and pension funds, recorded on a daily basis for a given trade execution date. Panel A presents contracts with maturities of 10 years or more, while Panel B displays contracts with maturities of 3 years or less. The dataset spans from January 2, 2019, to February 10, 2023, at the daily frequency. SOURCE: DTCC Trade Repository.

predominant investor type in each segment.

3 A model of the market and the identification problem

This section formalizes the demand and supply dynamics for insurance against inflation and characterizes the underlying determinants of inflation swap prices. The objective is not to provide a comprehensive structural model for empirical estimation, but rather to clarify key concepts, catalog relevant shocks, and precisely articulate the identification assumptions.

The model, motivated by the empirical facts, has two primary characteristics. First, it is a portfolio choice model, rather than a broker-dealer model, given the persistent large net positions held by dealer banks in this market. Second, it features two market segments, for short and long maturities, where inflation swaps are demanded by two distinct agents, hedge funds and pension funds, respectively, with dealer banks supplying both segments.

3.1 Each institution's problem

There are many traders working on behalf of institutions indexed by i that are grouped into three types: pension funds (f), hedge funds (h), and dealer banks (b). The set of institutions in each sector are $\Theta_f, \Theta_h, \Theta_b$, respectively, with measure given by the operator $|.|$. For exposition, we present the problem from the perspective of a trader acting on behalf of a bank indexed by b, i. Unless explicitly stated, traders acting on behalf of other institutions behave in the same way.

One period is a trading day. At the start of the day, the trader can buy (or sell) an inflation swap contract for a long maturity at a fixed price p that will pay off π at the end of the day. This payoff is not the realization of inflation, which will only be known years later, but rather the updated value on what future inflation is expected to be. The same applies to short maturity inflation swaps, which cost P and pay Π.

The trader also invests in a market asset that costs s and pays d. This could be a portfolio of many other assets, but we will consider solely their sum $e_{b,i}$ and treat the price s as exogenous, to focus on the demand for inflation risk.

Finally, each institution is exposed to background risk $y_{b,i}$ that cannot be traded due to incomplete markets, and which may be correlated with inflation. For a dealer bank, this could represent changes in the present value of other business lines, while for a pension fund it may reflect changes in the net present value of its payments to pension holders.

Combining all and normalizing the daily real safe rate to 1, the budget constraint of a dealer bank linking initial wealth $a_{b,i}$ to end-of-day wealth $a'_{b,i}$ is:

$$a'_{b,i} = a_{b,i} + (\pi - p)q_{b,i} + (\Pi - P)Q_{b,i} + (d - s)e_{b,i} + y_{b,i}\,, \tag{1}$$

where the net positions of long maturity and short maturity swaps are $q_{b,i}$ and $Q_{b,i}$, respectively.

Traders acting on behalf of hedge funds and pension funds have access to a similar set of assets. However, we impose one key restriction that reflects the facts in the previous section:

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

The goal of the trader is to maximize their institution's utility from terminal wealth:

$$\mathbb{E}_{b,i}\left[U(a'_{b,i})\right] , \tag{2}$$

according to an increasing concave function $U(.)$. Each institution has individual beliefs captured in the expectations operator $\mathbb{E}_{b,i}(.)$. We model disagreement over expected inflation as:

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

The parameters $\mu_{b,i}$ capture the heterogeneity in long maturity inflation expectations (with $\mu_{f,i}$ the equivalent for pension funds) and π^e is expected inflation, defined as the average across beliefs of institutions active in the long maturity market segment. The same applies for the short maturities with expected inflation Π^e and parameters $M_{b,i}$ and $M_{h,i}$ reflecting disagreement, so that institutions may disagree in different directions about short maturity and long maturity expected inflation.

Finally, there are capacity constraints on each institution's ability to take on inflation risk, capturing regulations, balance-sheet constraints, or investment mandates. We model these quite generally through two continuous functions:

$$G^L_b(q_{b,i}, Q_{b,i}, z_{b,i}) \geq 0 \quad \text{and} \quad G^S_b(Q_{b,i}, q_{b,i}, z_{b,i}) \geq 0\,, \tag{4}$$

that measure the proximity of the dealer bank to the capacity limits in its long maturity and short maturity trades, respectively. Each dealer faces an exogenous institution-specific shifter in the tightness of these financial constraints $z_{b,i}$.

We use asterisks to denote equilibrium values. Then, $\lambda^{L,*}_{b,i}$ is the Lagrange multiplier associated with the long maturity constraint at the optimal choice $q^*_{b,i}$, and $g^{L,*}_{b,i} \equiv \partial G^L_b(q^*_{b,i}, Q^*_{b,i}, z_{b,i})/\partial q_{b,i}$ for brevity, omitting the function's arguments. The terms $\lambda^{S,*}_{b,i}$ and $g^{S,*}_{b,i}$ are defined symmetrically. Pension funds also face capacity constraints in the long maturity segment of the market where they operate, with $\lambda^*_{f,i}$, $g^*_{f,i}$, $z_{f,i}$ as the Lagrange multiplier, the slope of a pension fund's capacity constraint at the optimal choice, and the exogenous shifter, respectively. The same applies to hedge funds in the short market with terms $\lambda^*_{h,i}$, $g^*_{h,i}$, and $z_{h,i}$.

3.2 Functional form assumptions

Since we will ultimately estimate linear regressions for the local slopes of demand and supply curves, we make functional assumptions that lead to a linear system from the portfolio choice.

This amounts to assuming a mean-variance optimization problem.

First, we assume preferences in the CARA class:

$$U(.) = -\exp\left(-\widetilde{\gamma}_{b,i} a'_{b,i}\right) . \tag{5}$$

The Arrow-Pratt absolute risk aversion coefficient is $\widetilde{\gamma}_{b,i} = \gamma_{b,i}/a_{b,i}$, where $\gamma_{b,i}$ is a fixed parameter, so that portfolio shares are independent of initial wealth.

Second, we assume that all the institutions believe that returns are normally distributed. While expected inflation is heterogeneous, we simplify the expectations of asset payoffs and background risk by assuming that all institutions agree that $\mathbb{E}_{b,i}[d] = \theta_d$ and $\mathbb{E}_{b,i}[y_{b,i}] = 0$, so that we focus on inflation.[9] The variances of the three exogenous random variables are σ^2_π, σ^2_d, and $\sigma^2_{y_{b,i}}$, while the covariances of expected inflation with market returns and background risk are $\sigma_{\pi,d}$ and $\sigma_{\pi,y_{b,i}}$, respectively, and their associated correlations are $\rho_{\pi,d}$ and $\rho_{\pi,y_{b,i}}$. Again to focus on inflation risk, we assume that $\sigma_{d,y_{b,i}} = 0$, or that background risk does not covary with market returns.

A final assumption is that inflation at different horizons covaries with market returns following a one-factor structure: $\rho_{\pi,\Pi} = \rho_{\pi,d}\rho_{\Pi,d}$. This simplifies the analysis by preventing inflation risk premia in one segment depending on banks' expectations in the other segment. In the data, the first factor explains 95% of the cross-sectional variance of the UK inflation swap curve at maturities from 1 to 15 years.

3.3 Individual asset holdings

Proposition 1, proven in Appendix B.1, states the asset demand function that solves b,i's optimization problem in the long maturity segment of the market (the equivalent expression for the short maturity segment is in the Appendix):

Proposition 1. *Given market prices p^* and s, a dealer bank's optimal demand for long maturity inflation swaps scaled by size is given by:*

$$\frac{q^*_{b,i}}{a_{b,i}} = \underbrace{\frac{\mu_{b,i}\pi^e - p^*}{\gamma_{b,i}\sigma^2_\pi(1-\rho^2_{\pi,d})}}_{\text{price and beliefs}} - \underbrace{\left(\frac{\sigma_d}{\sigma_\pi}\right)\left[\frac{\theta_d - s}{\gamma_{b,i}\sigma^2_d(1-\rho^2_{\pi,d})}\right]\rho_{\pi,d}}_{\text{hedging demand}} - \underbrace{\left[\frac{1}{(1-\rho^2_{\pi,d})\sigma^2_\pi}\right]\left(\frac{\sigma_{\pi,y_{b,i}}}{a_{b,i}} + \frac{\lambda^{L,*}_{b,i} g^*_{b,i}}{\gamma_{b,i}}\right)}_{\text{frictions}} \tag{6}$$

Demand for inflation swaps scales with the size of the institution 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 higher inflation, 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, so the slope of the individual trader's demand curve in quantity-price space will be higher the more risk averse it is.

[9] The message of the propositions is the same if these are institution-specific, but the algebra is more involved.

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

The third term captures the two frictions within the model. The first friction arises from the covariance of expected inflation with background risk, which when higher positive diminishes the demand for inflation protection due to the natural hedge provided by this income. For instance, a bank with a deposit franchise may benefit from rising nominal rates during periods of high inflation, which enhances earnings from its other business lines, leading to a positive $\sigma_{\pi,y_{b,i}}$. Conversely, for a pension fund with inflation-linked liabilities, this covariance is likely negative, creating an inherent demand for inflation protection. The second friction involves binding capacity constraints that reduce demand relative to the trader's desired level, constrained by regulatory, internal governance, or financial limitations. These capacity constraints are contingent on the equilibrium quantity supplied, thereby influencing the slope of the demand curve in the price-quantity space.

Similar expressions for hedge funds and pension funds are presented in Appendix B.2.

3.4 Demand, supply, and market clearing

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

$$q^* \equiv \sum_{i \in \Theta_f} q^*_{f,i} = -\sum_{i \in \Theta_b} q^*_{b,i} \quad \text{and} \quad Q^* \equiv \sum_{i \in \Theta_h} Q^*_{h,i} = -\sum_{i \in \Theta_b} Q^*_{b,i}\,, \tag{7}$$

so that $q^* > 0$ means that pension funds have net positive notional holdings, and we refer to them as demand, which then equals the supply by dealer banks in the long maturity segment. The same applies to Q^* in the short maturity segment where hedge funds and dealer banks meet.

Through the lens of our model, the fact that $q^* > 0$ can be explained by several forces. The most plausible is that pension funds are more exposed to background risk such that their net worth covaries more negatively with inflation: $\sigma_{\pi,y_{f,i}} < \sigma_{\pi,y_{b,i}}$. As described, many funds have liabilities in real terms through index-linked pension commitments to their members, which produces a negative covariance between net worth and inflation that is not present for banks.

Alternative explanations could be that pension funds systematically expect higher inflation $\mu_{f,i} > \mu_{b,i}$ or that banks face tighter capacity or regulatory constraints when taking long inflation positions: $\lambda^{L,*}_{b,i} g^{L,*}_{b,i} > \lambda^*_{f,i} g^*{}_{f,i}$. Differences in risk preferences show up in the slopes of supply and demand, and would lead to $q^* > 0$ if pension funds were more risk averse than banks, $\gamma_{f,i} > \gamma_{b,i}$, and $\rho_{\pi,d} > 0$ or $\pi^e < p^*$. This would be the case if inflation comoves positively with market returns.

3.5 The frictionless swap price and risk neutral expected inflation

Start with the case where 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 trading constraints do not bind for any agent in either market segment, $\lambda^{L,*}_{b,i} = \lambda^{*}_{f,i} = \lambda^{S,*}_{b,i} = \lambda^{*}_{h,i} = 0$. Combining equations (6) and (7), Appendix B.3 solves for this frictionless, counterfactual equilibrium price.

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

$$\widetilde{p} = \underbrace{\left[\frac{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i}\mu_{f,i} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i}\mu_{b,i}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i}}\right]}_{\equiv\Lambda,\ \textit{risk-adjusted size-weighted average of beliefs}} \underbrace{\pi^e}_{\textit{expected inflation}} - \underbrace{\left(\frac{\theta_d - s}{\sigma_d\sigma_\pi}\right)\rho_{\pi,d}}_{\textit{risk premium}}. \tag{8}$$

These two components of equilibrium prices map into the objects that policymakers care about and that models focus on. The first is expected inflation. The coefficient on π^e, which we denote Λ, is a weighted function of each institution's beliefs and their size-adjusted risk preferences $\widetilde{\gamma}_{f,i}$ and $\widetilde{\gamma}_{b,i}$. How the heterogeneity in beliefs, preferences, and wealth affects prices depends on the correlation amongst them. If beliefs are independent of risk aversion at the institution level, then $\Lambda = 1$, and the frictionless price rises one for one with π^e. This aligns with the typical interpretation that movements in swap prices provide a direct measure of changes in expected inflation.

The second component of the price is compensation for risk, arising because of the correlation of inflation with the return on the other assets that pension funds and dealer banks invest in $\rho_{\pi,d}$.[10] Policymakers and academics care about this term because it captures both the variance of inflation, and how it co-moves with real outcomes.

Combining the two, the frictionless price is a measure of risk-neutral expected inflation. It varies over time with changes in the macroeconomy, monetary policy, or the regime that drives inflation. We denote shocks to risk-neutral expected inflation by ε_π and refer to them as inflation shocks. Importantly, they affect the price because they shift both demand and supply in the market. Changes to risk-neutral expectations of inflation affect the common fundamental reasons that drive the institutions to participate in the market in the first place.

The same applies to the short maturity segment (see Appendix B.3 for the equivalent expression for the frictionless short maturity price). We assume there is a single shock to inflation expectations that affects both segments, but the relative sign and scale of the movements at each maturity is arbitrary, so the same ε_π that shifts $\widetilde{p}$ can shift $\widetilde{P}$ in any direction. This means that an

[10] With heterogeneity in individual beliefs about this covariance, the inflation risk premium would included a weighted average of the heterogeneous covariances.

inflation shock will shift supply and demand in both segments simultaneously.[11]

3.6 Friction-driven price movements

Because of the two general frictions—incomplete markets and trading constraints—the actual price deviates from its frictionless value, as derived in Appendix B.3.

Lemma 2. *The price of a long maturity swap is:*

$$p^* = \widetilde{p} - \underbrace{\frac{\sum_{i\in\Theta_f}\left\{\sigma_{\pi,y_{f,i}} + \frac{\lambda^*_{f,i} g^*_{f,i}}{\widetilde{\gamma}_{f,i}}\right\}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i}}}_{\textit{frictional demand from pension funds}} - \underbrace{\frac{\sum_{i\in\Theta_b}\left\{\sigma_{\pi,y_{b,i}} + \frac{\lambda^{L,*}_{b,i} g^{L,*}_{b,i}}{\widetilde{\gamma}_{b,i}}\right\}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i} + \sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i}}}_{\textit{frictional supply from dealer banks}}. \tag{9}$$

Changes in the ability of pension funds to trade inflation swaps, $z_{f,i}$, can tighten capacity constraints and so move $\lambda^*_{f,i}$. Changes in $\sigma_{\pi,y_{f,i}}$, capturing how the income flow of pension funds covaries with inflation, will likewise change demand for inflation protection. The aggregate demand curve combines these two components, and its impact on the equilibrium price then depends on the slope of demand and supply, captured in the second term on the right-hand side of equation (9). We denote shocks to these frictions that drive pension fund demand by ε_f.

The third term on the right-hand side of equation (9) likewise shows that the impact of frictions on dealer banks' willingness to supply inflation swaps affects equilibrium prices. These come from shifts in banks' trading capacity, $z_{b,i}$, or in the income risk, $\sigma_{\pi,y_{b,i}}$. These shift the supply curve, and we denote shocks to them by ε_b.

A similar expression applies to the short maturity price P. The same shocks to the frictions afflicting dealer banks, ε_b, affect supply in the short maturity segment, together with the shocks to the demand from hedge funds, which we denote by ε_h. Demand shocks in one segment spill over to the other because they affect the dealer's ability to supply swaps in that segment. This is captured by the $g^{L,*}_{b,i}$ term above, since a shock ε_h will affect the quantity supplied in the short maturity segment Q^* and this shifts the quantity constraint in the long maturity segment $G^L_b(q^*_{b,i}, Q^*_{b,i}, z_{b,i})$, and so its derivative $g^{L,*}_{b,i}$ and the long maturity price p^*.

3.7 The identification problem

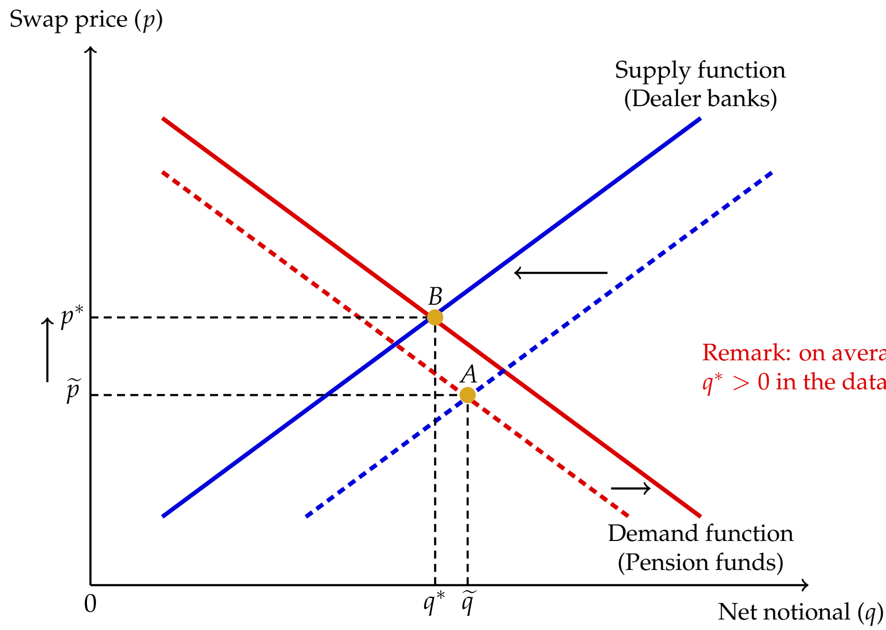
Figure 6 graphically displays the two market segments under different circumstances.

Panel (a) describes the long maturity segment where pension funds are net buyers of inflation protection from dealer banks. The frictionless equilibrium at point A reflects expected inflation

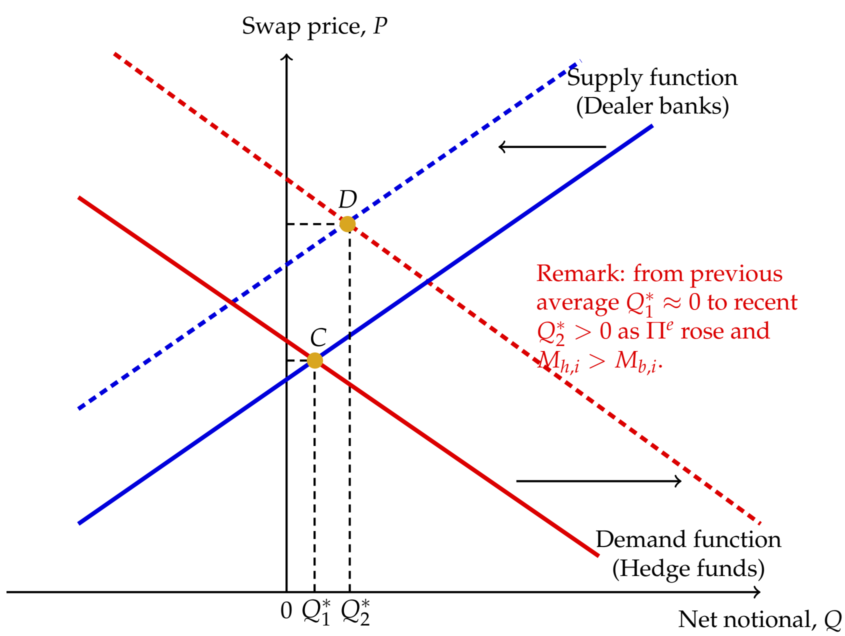
[11]The assumption that there is one inflation shock is not particularly restrictive; the forecast revision in the long and short segments from one period to the next can move in any arbitrary direction. For instance, this allows underlying inflation to follow any arbitrary ARIMA process. This is again consistent with the empirical observation that a single factor explains the vast majority of the cross-maturity variation in the inflation swap curve.

Figure 6 FRICTIONLESS AND OBSERVED PRICES AS OUTCOMES OF SUPPLY AND DEMAND

(a) The long maturity segment and the effect of shocks to supply (ε_b) and demand (ε_f)



(b) The short maturity segment and the response to an inflation shock (ε_π)



NOTE: Panel (a) describes the long maturity segment where pension funds trade against dealers and are net buyers of inflation swaps ($q > 0$). The point A is the frictionless equilibrium, where pension funds and dealers have no trading constraints and background risk, and the price is $\widetilde{p}$. Correspondingly, the blue dashed line is the dealer supply curve in the frictionless case while the red dashed line indicates the frictionless pension fund demand curve. Introducing dealer supply constraints shifts the blue line up-left, resulting in a solid dashed line. Frictions affecting pension fund demand are assumed to shift the red line up-right. The resulting frictional equilibrium is point B, with corresponding observed price p^*. In other circumstances, the supply/demand frictions may shift supply and demand differently, resulting in $p^* < \widetilde{p}$. Panel (b) describes the short maturity segment where hedge funds trade against dealers. The initial equilibrium is denoted C, with solid demand and supply curves and a small net notional position, Q_1^*. The panel shows the response to a positive inflation shock ($\varepsilon_\pi > 0$), resulting in the dashed supply and demand curves and a new equilibrium D. The rise in equilibrium prices reflects a rise in expected inflation.

and compensation for its perceived volatility, captured by the dashed counterfactual supply and demand functions where pension funds and dealer banks are not afflicted by the frictions described above.

In the example in the figure, dealer banks' supply constraints reduce their supply left-up to the actual supply in the blue solid line, while pension funds' income risk raises their demand right-up to the actual demand curve in the red solid line. The actual equilibrium is at point B, with a swap price p^* above risk-neutral expected inflation $\widetilde{p}$. A shock to the dealer banks that tightens their constraints $\varepsilon_b > 0$ or a shock to pension funds that raises their desire for inflation protection $\varepsilon_f > 0$ would further shift up supply or demand, respectively, and raise the actual price even though expected inflation is unchanged.

Panel (b) describes an initial equilibrium C in the short maturity market segment where hedge funds trade against dealers with a resulting small net notional position Q_1^*. If banks and hedge funds have similar beliefs, hedging demand, and background risk, the model would predict that equilibrium net holdings are on average close to zero, as is the case in the data. Following a positive inflation shock ($\varepsilon_\pi > 0$), we get the dashed supply and demand curves and a new equilibrium D. The new price signals a rise in expected inflation. As drawn, $Q_2^* > Q_1^*$, which would be the case if on average (adjusting for size and risk aversion) $M_{h,i} > M_{b,i}$, such that hedge funds react more to a shock to expected inflation than dealers.

The identification problem is that a researcher observing only the rise in prices p^* and P^* would be unable to determine if they rose due to an increase in risk-neutral expected inflation, like in Panel (b), or due to frictions, like in Panel (a). With four shocks driving the market—one to inflation, and three demand/supply frictions—it is difficult to disentangle the individual effects.

However, we are able to utilize daily data on prices *and* quantities. The data in the column vector $\mathbf{Y} = (Q, P, q, p)'$ could identify the shocks in the vector $\boldsymbol{\varepsilon} = (\varepsilon_h, \varepsilon_f, \varepsilon_b, \varepsilon_\pi)'$.[12] One example of how this helps comes from the figure: the inflation shock led to an increase in the net notional in the short maturity market segment, while the frictional shocks led to a fall in the quantity in the long maturity market segment.

However, quantity data alone is not enough. For instance, while the rise in prices and quantities in the bottom panel matches what we observe in the data, the post-pandemic recovery may have loosened trading restrictions for both, and relatively more so for hedge funds, which would match the figure just as well. Formally:

$$\mathbf{Y} = \boldsymbol{\Psi}\boldsymbol{\varepsilon}\,, \tag{10}$$

and we want to pin down the elements of the 4×4 matrix $\boldsymbol{\Psi}$ to fully identify the system, or at

[12] To be more precise, our data is: (i) for q, net purchases from dealers of UK RPI inflation swaps with initial time-to-maturity of 10 years or more; (ii) for p a weighted-average daily price of UK RPI zero coupon inflation swaps with initial time-to-maturity of 10 years or more, where the weights are constructed using the sample average gross notionals traded at each maturity, and adjusted for indexation lags; (iii) for Q, the net purchases of swaps with initial time-to-maturity of 3 years or less; and (iv) for P, the volume-weighted average daily price of UK RPI zero coupon inflation swaps of maturity less than 3 years, again adjusted for indexation lags.

least four elements in its inverse to partially identify the shock to expected inflation ε_π.

4 Three identification strategies

Our sample spans from January 2, 2019, to February 10, 2023, encompassing a period characterized by highly volatile inflation and expected inflation. This dataset includes 1,078 daily observations, covering 210 pension funds, 30 hedge funds, and 16 dealer banks (13 in the short market). In this section, we detail how we leverage three distinct sources of variation in the data—over time, across institutions, and at daily frequency—to achieve identification through three different strategies. For brevity, we first describe a static case and subsequently incorporate dynamics when describing the estimation strategy.

4.1 Strategy 1: heteroskedasticity across time

Inflation news tends to be released in a lumpy manner. Although the underlying shocks may be smooth over time, traders gain the most significant insights on the dates when official data on prices is released. In our sample, the volatility of inflation swap prices and the quantities traded is markedly higher on these dates.

While it is possible that supply and demand shocks also become more volatile during these periods, it is plausible to attribute the observed spike in variance on data release dates to the revelation of inflation shocks. This forms our first identifying assumption. Formally:

Assumption B1. *(Heteroskedasticity in the inflation shock at known dates.) Let $\mathbf{\Sigma}_H$ denote the diagonal variance-covariance matrix of the shocks $\boldsymbol{\varepsilon}$ at data release dates, and $\mathbf{\Sigma}_L$ the one at other dates. Assume that: (i) the largest diagonal of $\Sigma_H \Sigma_L^{-1}$ is greater than one, unique, and corresponds to the ratio of variances of the inflation shock between release and non-release dates; and (ii) Ψ does not change between release and non-release dates.*

The first part of the assumption states that the relative variance of the shocks changes between two known set of dates due to an increase in the variance of the inflation shock. The second part states that the propagation from shocks to the observables does not change. Combined, these assumptions identify (up to sign and scale) the column of Ψ associated with the inflation shock. Specifically, if $\mathbf{\Omega}_H$ denotes the variance of $\mathbf{Y}$'s unpredictable component on release days, and $\mathbf{\Omega}_L$ is the equivalent for the other dates, then the relevant column of Ψ is the left eigenvector of $\mathbf{\Omega}_H \mathbf{\Omega}_L{}^{-1}$ corresponding to the largest eigenvalue (Lewis, 2024). Note that since, *a priori*, the relative variances of demand and supply shocks may rise or fall at those dates, this strategy only allows for the identification of ε_π and the associated column of Ψ, not the complete system.

In our sample, we have 49 monthly dates when the data on UK RPI inflation was released. Two more dates had a large impact on expected inflation: 8^{th} September 2022 when former Prime Minister Truss announced a cap on energy prices in the UK parliament, and 23^{rd} September 2022 when former Chancellor Kwarteng announced the "Mini-Budget"—both announcements dramatically changed the properties of measured inflation. This gives a total of 51 of dates.[13] In the data, the largest eigenvalue of $\Omega_H\Omega_L^{-1}$ is 1.43, so inflation shocks have significantly higher variance on release dates.

4.2 Strategy 2: instrumental variables using cross-institutional granularity

Asset demand for an individual institution in proposition 1 is the sum of common drivers and idiosyncratic institution-level frictions. In turn, the solution for observed prices in lemma 2 depends on risk-neutral expected inflation and aggregate frictions. Combining the two, we can write observed net notionals of an institution as a reduced-form factor model.

For pension fund f,i, let $\mathbf{F}$ denote unobserved common factors, $\boldsymbol{\omega}_{f,i}$ denote the institution-specific factor loadings, and $\tilde{\varepsilon}_{f,i}$ are idiosyncratic shocks to frictions (all appended with time subscripts). Then:

$$\frac{q_{f,i,t}}{a_{f,i,t}} = \boldsymbol{\omega}'_{f,i}\mathbf{F}_t + \widetilde{\varepsilon}_{f,i,t}\,. \tag{11}$$

Appendix C.1 shows the exact expressions for each of these components.

Given estimates of this model, one can calculate a wealth-weighted average of the residuals across institutions:

$$GIV_{f,t} = \sum_{i\in\Theta_f} a_{f,i,t}\tilde{\varepsilon}_{f,i,t}\,. \tag{12}$$

Since $\mathbf{F}_t$ captures the aggregate drivers of demand and supply, this new variable is by construction independent of inflation shocks, $\mathbb{E}(GIV_{f,t}\varepsilon_{\pi,t}) = 0$, and of the shocks to supply by banks, $\mathbb{E}(GIV_{f,t}\varepsilon_{b,t}) = 0$. Assumption A ensures that $\mathbb{E}(GIV_{f,t}\varepsilon_{h,t}) = 0$. Therefore, the exclusion restriction for $GIV_{f,t}$ to serve as an instrument for $\varepsilon_{f,t}$ is satisfied.

If the average was unweighted, it would equal zero by construction, since the estimated values of $\mathbf{F}_t$ always span the period mean. The instrument $GIV_{f,t}$ would not be relevant, as it would be uncorrelated with $\varepsilon_{f,t}$. However, some institutions are significantly larger than others in the data. Individual shocks to their demand function do not average out, but drive the aggregate demand in the market. Pension funds' daily gross notional positions in long maturity swaps are well described by Zipf's law, with an estimated power coefficient of -0.9. The same applies to dealer banks and hedge funds; see Appendix C.2.

This provides the source of variation to satisfy the relevance condition. More precisely, the assumption of granularity that replaces assumption B1 in this strategy is:

[13] In the supplemental Appendix, as a robustness check, we augment the set of high variance dates to include the 33 days in the sample when there was a monetary policy decision. The results are nearly unchanged, but since the evidence for heteroskedasticity is weaker with these, we only keep the RPI release dates as the benchmark specification.

Assumption B2. *(Granularity of the institutions.) The data on asset positions $a_{f,i,t}$, $a_{h,i,t}$ and $a_{b,i,t}$ are granular in that:*

$$\mathbb{E}(GIV_{f,t}\varepsilon_{f,t}) \neq 0 \;\; and \;\; \mathbb{E}(GIV_{b,t}\varepsilon_{b,t}) \neq 0 \;\; and \;\; \mathbb{E}(GIV_{h,t}\varepsilon_{h,t}) \neq 0\,. \tag{13}$$

Appendix C.3 shows that with this assumption, the three instruments for the three frictional shocks are valid. To give a sense of relevance, the F-statistics for these three instruments in a frequentist model are: 72.4 for $GIV_{f,t}$, 22.3 for $GIV_{h,t}$ and 43.5 for $GIV_{b,t}$. Since the system has four shocks and we have three instruments, the inflation shock is also identified together with the full matrix Ψ; see Appendix C.4.[14]

We implement this strategy by estimating equation (11) using an interactive fixed effects model (Bai, 2009). Our stylized model suggests a two-factor structure, but it ignores the potential for groups within each sector (e.g., public versus private pension funds) or connections between clients and dealers. We therefore allow the data to determine the appropriate number of factors for each sector. Starting with a candidate number of factors derived from the information criterion of Bai and Ng (2002), we verify whether the resulting point estimates align with supply and demand shocks (i.e., whether prices and quantities move in opposite or the same direction within the relevant market segment). We incrementally increase the number of factors until this condition is met, ensuring that the final selection remains robust even when additional factors are included. Appendix C.5 describes the procedure in more detail.

We do not have data on total assets, but we can proxy $a_{f,i,t}$ by using gross notional positions in the long maturity segment for pension funds, the short maturity segment for hedge funds, and the sum of both for dealer banks. Measurement error in $a_{f,i,t}$ is not problematic, as long as the resulting instruments remain relevant.

4.3 Strategy 3: segmentation and heterogeneity in high-frequency reactivity

Within a trading day and within a dealer bank, it is plausible that the ability of traders to take positions in the short maturity market segment is not constrained by the position taken in the long maturity segment. In many institutions, the desks that sell long maturity swaps to pension funds and short maturity swaps to hedge funds are often separated and staffed by different traders. Trading books are closed at the end of day so that the spillovers from positions in one segment—that may offset or constrain activity in the other—only manifest on the following day. Formally:

Assumption B3. *(Desk separation within the day.) The dealers' capacity constraints are independent of each other: $\partial G_b^S(\cdot,\cdot)/\partial q_{b,i} = 0$ and $\partial G_b^L(\cdot,\cdot)/\partial Q_{b,i} = 0$ so that they are:*

$$G_b^S(Q_{b,i}, z_{b,i}) \geq 0 \;\; and \;\; G_b^L(q_{b,i}, z_{b,i}) \geq 0\,. \tag{14}$$

[14] Of course, the shocks may not be precisely identified. In our application, while this strategy gives sharp estimates for $\varepsilon_{f,t}, \varepsilon_{h,t}$ and $\varepsilon_{\pi,t}$ and their corresponding impulse responses, the supply shock $\varepsilon_{b,t}$ is not precisely estimated.

Second, also within a day, it is plausible to expect that different types of institution react faster to news about inflation. In the long maturity market segment, dealers are likely more informed (or attentive/reactive) than pension funds. After all, dealers that trade inflation risk see all sides of the market, so they will have more precise posterior information about inflation. In contrast, in the short maturity segment, informed hedge funds are likely more reactive to inflation news than dealer banks. Formally:

Assumption B4. *(Differential reactiveness to news about inflation.) Dealer banks respond more to long maturity expected inflation than pension funds but less to short maturity expected inflation than hedge funds:*

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

$$\frac{\sum_{i\in\Theta_h}\widetilde{\gamma}_{h,i}^{-1}M_{h,i}}{\sum_{i\in\Theta_h}\widetilde{\gamma}_{h,i}^{-1}+\sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}} > \frac{\sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}M_{b,i}}{\sum_{i\in\Theta_f}\widetilde{\gamma}_{h,i}^{-1}+\sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}}. \tag{16}$$

These two assumptions imply the following sign and zero restrictions on the response of observables to shocks within a day that fully set identify the system:

$$\boldsymbol{\Psi} = \begin{pmatrix} + & 0 & - & + \\ + & 0 & + & + \\ 0 & + & - & - \\ 0 & + & + & + \end{pmatrix}. \tag{17}$$

Starting with the first two columns, shocks to the frictions affecting pension funds and hedge funds shift the demand curves in the long and short maturity segments, respectively. Therefore, in the market segment where the funds are active, they drive prices and quantities in the same direction, hence the positive signs.

Assumptions A and B3 collectively imply that within a trading day, there is no spillover effect from the quantity demanded in one segment into dealer trading constraints in the other segment. Consequently, shocks to pension fund demand have zero impact on prices or quantities in the short horizon segment, and similarly, hedge fund demand shocks have no effect on the long horizon segment. This rationale explains the zeros in the first two columns.

A dealer supply shock results in quantities and prices moving in opposite directions across both market segments simultaneously, which justifies the signs in the third column.

Assumption B4 justifies the fourth column. Following an inflation shock, the upward shift in the supply function outweighs the shift in the demand function, causing p to rise and q to fall. Conversely, in the short horizon market segment, the upward shift in the demand function prevails, leading P and Q to move in the same direction.

These assumptions leverage the availability of high-frequency data on quantities and prices. It is unlikely that desks remain separated over a month or even a week, and any informational differences between banks, hedge funds, and pension funds may dissipate within a few days (as we will observe). However, at daily frequency, these assumptions are plausible and deliver identification.

4.4 Dynamics and implementation

Thus far, we have examined the model and identification within a static framework. If markets were efficient and shocks dissipated within a single day, this approach would be sufficient.

In practice, information may diffuse slowly, and markets may only gradually incorporate it into prices. Introducing dynamics into the model has minimal impact, as we can simply reclassify the shocks as state variables that follow their own time series processes. For the empirical implementation, we incorporate dynamics by estimating a structural VAR with a deterministic constant $\mathbf{c}$ and three lags (selected with the Bayesian information criterion):

$$\mathbf{Y}_t = \mathbf{c} + \sum_{\ell=1}^{3} \mathbf{\Phi}_\ell \mathbf{Y}_{t-\ell} + \mathbf{u}_t \quad \text{and} \quad \mathbf{u}_t = \mathbf{\Psi} \boldsymbol{\varepsilon}_t \,. \tag{18}$$

We estimate the VAR using Bayesian methods with diffuse priors. For the first strategy, we use a variant of the sampler in Brunnermeier et al. (2021). For the second strategy, we use the GIV as proxy instrumental variables with a modified version of the sampler in Bahaj (2020). For the third strategy, we use the sampler from Arias et al. (2018).[15]

To compare the estimation results across the three identification strategies, we scale the impulse response functions such that the increase in the net notional position of hedge funds in the short horizon market segment, Q, equals \$1bn in response to either the inflation shock, hedge funds' demand shock, or dealers' supply shock. Similarly, pension funds' demand shock in the long horizon market segment is scaled to raise q by \$1bn.

5 Estimates of expected inflation

This section shows our main estimates of interest—the inflation shocks ε_π and the corresponding movements in risk-neutral expected inflation—and compares them with the observed prices.

5.1 The shocks across identification strategies

The correlation of the time series of inflation shocks identified by heteroskedasticity and by instrumental variables is 0.98; between heteroskedasticity and sign restrictions is 0.84; and between in-

[15]The construction of posterior distributions is detailed in the supplemental Appendix.

strumental variables and sign restrictions is 0.89.[16] Each of the three strategies leverages a distinct feature of the data: the time series dimension, variation across institutions, and high-frequency structure of the data. The remarkably high degree of correlation across the estimates reinforces the robustness of our inferences.

Having three identification strategies, we can also internally cross-verify the separate assumptions that underpin each of them. Namely, using the estimates of $\hat{\varepsilon}_{\pi,t}$ from one strategy, we can test whether the assumptions B1-B4 of the other strategies hold.

Starting with assumption B1 on heteroskedasticity, if we use the series of inflation shocks identified by granular instrumental variables, the variance of the inflation shock on release dates versus non-release dates is higher for 99% of draws.

Second, to investigate the exogeneity of our granular instruments, we calculate the three sample analogs of the exclusion restrictions $\frac{1}{T}\sum_{t=1}^{T} GIV_{\nu,t}\hat{\varepsilon}_{\pi,t}$ for $\nu \in \{f, h, b\}$ scaling both instrument and shock to have unit variance. Using the estimates of inflation shocks identified via sign restrictions, the values are -0.0068, 0.0089, and 0.031 for $\nu = f, h, b$, respectively. Using instead the estimate of the inflation shock identified via heteroskedasticity, they are $-0.061, 0.012, 0.056$, respectively. All are close to zero, supporting the exogeneity of the granular instrumental variables.

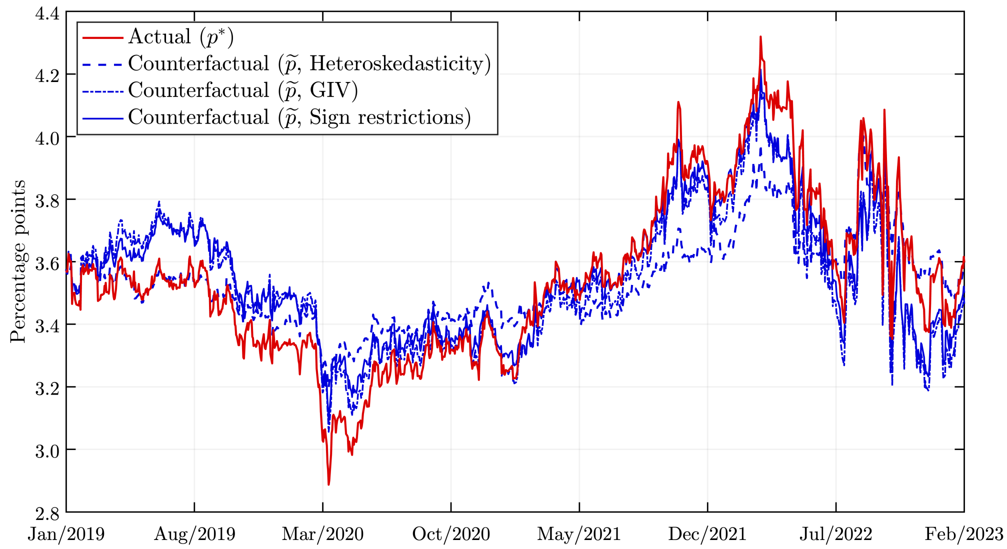
Third, we compute the impulse response functions to an inflation shock using both the heteroskedasticity strategy and the granular instrumental variables strategy, and verify that their signs align with the sign restriction strategy. The responses will be discussed in detail in Section 6.1. Prices increase on impact in both market segments, while the quantity traded rises in the short maturity segment and falls in the long maturity segment under the heteroskedasticity strategy, supporting Assumption B4. (The median estimate indicates an increase in quantities in the long maturity market in response to an inflation shock identified using instrumental variables, but the confidence bands are sufficiently wide to not reject the assumption of a negative response.) Regarding Assumption B3, under the granular instrumental variables specification, the 90% Bayesian credible intervals for both quantity and price responses in the short maturity market segment include zeros in response to a shift in the demand for long maturity inflation protection from pension funds. The same holds true in the long maturity segment in response to hedge fund demand shocks.

5.2 The evolution of risk-neutral expected inflation

Figure 7 shows, in red, the long maturity swap price p^* and, in blue, the counterfactual evolution of the swap price if there had only been shocks to risk-neutral expected inflation $\widetilde{p}$, for each of the three identification strategies. Since the level of the counterfactuals is not identified, we normalize their initial value to be the same as p^*, as is standard practice in historical decompositions.

[16] Each shock time series is constructed using the median of the parameter estimates from the sampler for each identification strategy.

Figure 7 SWAP PRICES AND THE CONTRIBUTION OF EXPECTED INFLATION AT LONG HORIZONS



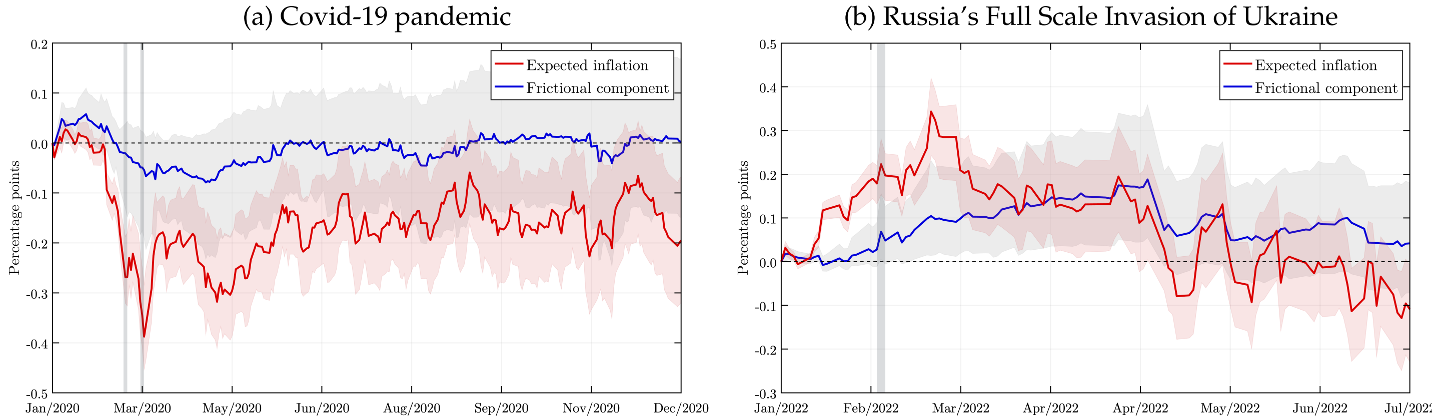
NOTE: The long maturity inflation swap breakeven rate, p^*, is shown in red. The counterfactuals, depicted in blue, represent the breakeven rate after subtracting the cumulative effects of the frictional shocks $\varepsilon_b, \varepsilon_h, \varepsilon_f$, as estimated by each of the three identification strategies.

The correlation between realized and counterfactual prices is quite high. Still, the two sometimes differed by 20-30 basis points, for instance in June-July of 2020, and in October-December of 2021, crucial times for monetary policy as it tried to respond to the pandemic lockdowns and to the impact of the impending invasion of Ukraine. These are significant differences that could change conclusions on whether markets expected inflation to deviate significantly from target.

During this specific sample, swap prices have overstated the fluctuations in long-horizon inflation expectations: when expected inflation fell, the swap price fell by more; when it rose, it rose by more. Our estimates suggest that the risk of inflation becoming unanchored was overstated because of frictional shocks.

To further investigate this pattern of overshooting, Figure 8 focuses on two specific episodes that led to significant movements in expected inflation and presents a price decomposition using the sign restriction strategy (see Appendix D.2 for the underlying price movements). The first episode was the onset of the Covid-19 pandemic, which triggered an immediate recession and deflation. The second episode was Russia's full-scale invasion of Ukraine, which caused a sharp rise in UK inflation due to disruptions in energy supply. Following both events, expected long-horizon inflation (in red) shifted by 20-40 basis points, mirroring the direction of actual inflation. Simultaneously, in both events, the frictional component of swap prices (in blue) moved in the same direction as expected inflation. Policymakers observing swap prices would have overstated the extent of the change in long-run expectations and potentially reacted too strongly to the

Figure 8 TWO EPISODES: EXPECTED INFLATION AND FRICTIONS



NOTE: Estimated cumulative contribution of inflation shocks ε_π and all frictional shocks $\varepsilon_b, \varepsilon_f, \varepsilon_h$ to the price series using the median estimate from the sampler of the estimates from the sign restriction strategy, with shaded areas representing 68% credible intervals.

shock.[17]

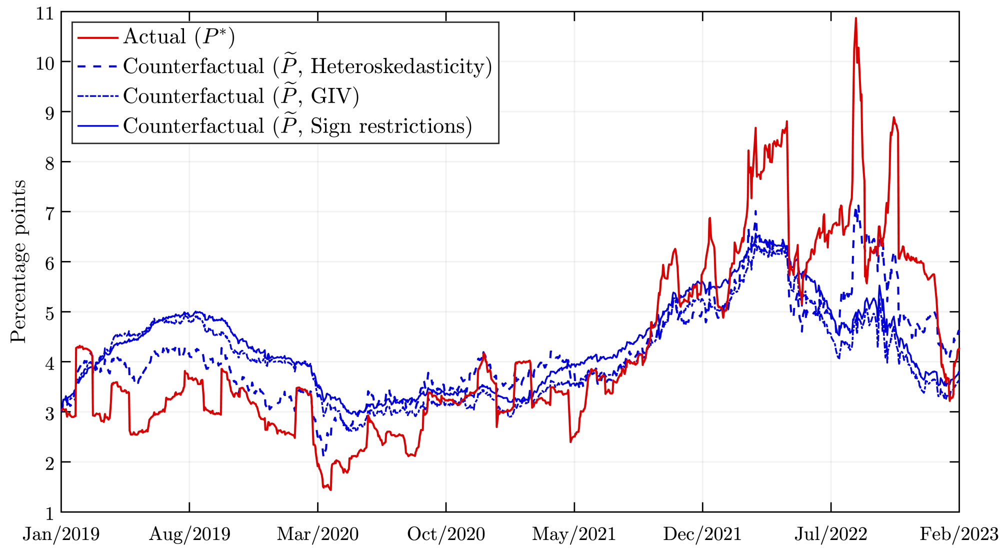
As a rule of thumb, on average over our sample, when the swap price rises by 100 basis points, the best estimate is that actual expected inflation increases by only 87.6 basis points (section 6.3 will discuss the full forecast error decomposition). During significant events, the overstatement is larger, with only approximately three-quarters of the observed change in long maturity swap prices corresponding to changes in expected inflation.

Figure 9 presents swap prices and the corresponding counterfactuals for short maturities. The deviations between the two series are both larger and more persistent. The series can diverge by more than 100 basis points for several weeks. On average, an unexpected 100 basis point increase in short maturity swap prices within a single day corresponds to only a 6.4 basis point change in expected inflation. Even when this unanticipated move in short-term prices accumulates over three months, the share accounted for by expected inflation rises to only 24.3 basis points.

This underscores the importance of caution for users of swap price data when measuring expected inflation at short horizons: it is essential to adjust for frictions.

[17]The LDI crisis in autumn 2022 provides a third episode of volatile prices, driven by a combination of a financial liquidity crisis and fiscal interventions. This is discussed separately in Appendix D.3.

Figure 9 RISK-NEUTRAL EXPECTED INFLATION AND SWAP PRICES AT SHORT HORIZONS



NOTE: The short maturity inflation swap breakeven rate, P^*, is shown in red. The counterfactuals, depicted in blue, represent the breakeven rate after subtracting the cumulative effects of the frictional shocks $\varepsilon_b, \varepsilon_h, \varepsilon_f$ estimated by each of the identification strategies.

6 Estimates of how the market shifts inflation risk

To estimate shocks to risk-neutral expected inflation, we identified supply and demand shocks in the inflation swap market. These are interesting in their own right as they offer a unique opportunity to study a large OTC derivatives market linked to a key macroeconomic variable, exhibiting striking market segmentation. This section discusses the impulse response functions, the estimated slopes of the supply and demand functions, and the relative importance of various frictions.

6.1 The speed of price adjustment to shocks

Figure 10 shows the impulse response to an inflation shock for the three identification strategies. The responses are similar and credible sets overlap.[18] The market for inflation swaps adjusts quickly to a shock. All impulse responses stabilize within one to three days and remain persistent thereafter. While quantities go back to a steady state after a shock, prices change persistently, consistent with the martingale property of a (weakly) informationally efficient financial market.

Because our main results are qualitatively similar across different strategies, we will report results using the sign restriction strategy in this section to conserve space. This strategy identifies the entire system and precisely estimates all four shocks. Relevant results from other strategies

[18]The exception is the quantity response in the long-horizon segment using the granular instrumental variable strategy, which is positive, whereas it is negative with the other two identification strategies. However, the uncertainty surrounding this estimate is high, and the negative responses from the other strategies fall within its credible interval.

are presented in the Appendix.

6.2 The slopes of demand and supply

An identified shock to the pension funds' demand curve in the long maturity market segment ε_f moves prices and quantities along the supply curve. Dividing the estimated impact on the price by the estimated impact on the quantity gives a local estimate of the slope of the supply curve by dealer banks in that segment of the market. Applying the same logic to each of the other frictional shocks identifies the demand and supply curves in both segments. Figure 11 presents the estimated supply and demand curves in both market segments alongside credible intervals.[19]

Comparing the demand curves across the two segments, the median estimate of the slope of hedge funds' demand for short maturity protection is -0.61 percentage points per \$1bn, whereas for pension funds' demand for long maturity protection, it is -0.52 percentage points per \$1bn. These units obscure the differing volatilities of prices: long maturity prices must decline by two standard deviations to induce a \$1bn increase in pension fund demand, while a 0.35 standard deviation change in short maturity prices results in an equivalent rise in hedge fund demand. Within the framework of our model, this suggests that pension funds are effectively more risk-averse or face steeper quantity constraints on trading compared to hedge funds, potentially due to having to align their exposures with their long-term liabilities.[20]

The more striking comparison lies in the slopes of the two supply functions by the dealers. In the long maturity segment, the supply curve is flat; conversely, in the short maturity segment, the supply curve is slightly steeper than the demand curve. The model would interpret this as $g^L_{b,i}$ being locally insensitive to $q_{b,i}$, while $g^S_{b,i}$ significantly varies with $Q_{b,i}$. That is, the long maturity supply curve is almost horizontal because dealers are able to trade more freely in the long maturity segment versus the short maturity segment.

Beyond the assumptions of the model, another potential explanation for the striking contrast in the slopes of supply curves by the dealers is that trading against informed traders, such as hedge funds, is subject to adverse selection, which is not present when trading against pension funds. Another alternative is that dealers have market power over pension funds and operate with variable markups. For example, if pension costs face a fixed cost of switching dealers, the optimal markup their existing dealer can charge will be declining in quantities, therefore flattening the effective supply curve. Determining which of these explanations drives the estimates is an exciting challenge for future research.[21]

[19] The corresponding impulse response functions are presented in the supplemental Appendix. Because, as was the case with inflation shocks in Figure 10, quantities revert quickly back to zero, calculating the slope using the impact response is appropriate.

[20] Bretscher et al. (2021) and Jansen (2025) also find that pension funds exhibit relatively lower responsiveness to price changes in other markets.

[21] We present the equivalent demand and supply curves for the GIV strategy in Appendix D.1 confirming the flat supply curve in the long maturity segment versus the short maturity segment. The slopes of the demand curves are less precisely estimated using this strategy however.

6.3 The drivers of inflation swap prices

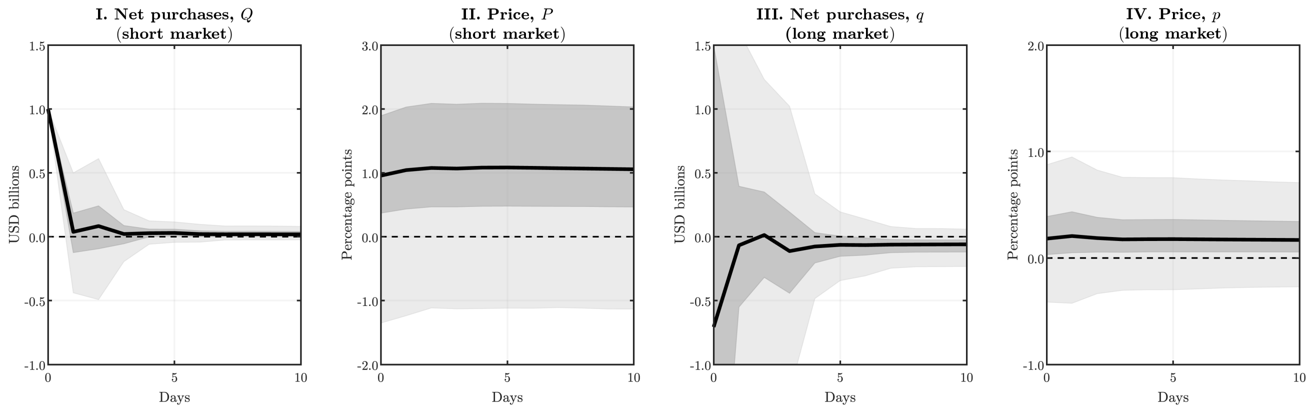
Figure 12 presents the forecast error variance decompositions for the different shocks across various forecast horizons. Due to the flat supply curve, frictional shocks to demand are irrelevant for the long maturity swap prices. Approximately three-quarters of the price variation is attributed to shocks to risk-neutral expected inflation, with the remaining quarter explained by frictions shifting the supply by dealer banks. This explains the close alignment of swap prices with expected inflation observed in Figure 7.

In the short maturity segment of the market, where both supply and demand curves are sloped, frictional shocks dominate expected inflation in influencing prices. Inflation shocks only significantly impact unexpected movements in short-maturities swap prices at longer forecast horizons, accounting for roughly 30% of the variation over 200 days. Again, this aligns with our comparison of swap prices and expected inflation in Figure 9, which showed notable comovement only at low frequencies. While the literature primarily focuses on constraints faced by banks, our finding that demand shocks to hedge funds are also crucial for prices over longer forecast horizons suggests that future research should explore how the frictions and risks faced by such institutions correlate with inflation.

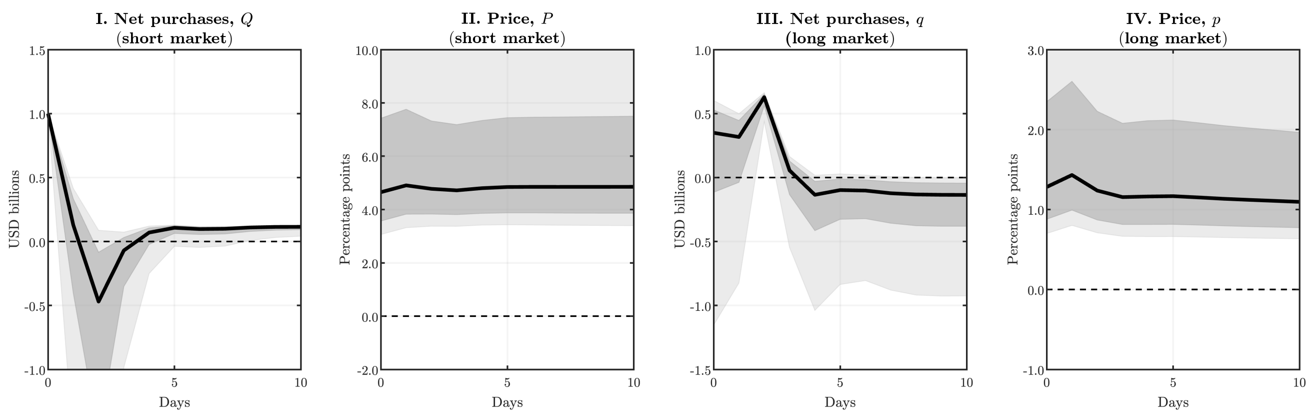
Turning to quantities in each segment of the market, frictional shocks to demand by pension funds account for almost all of the variation in the long maturity segment. The flat supply curve implies that demand shocks affect quantities rather than prices. In the short maturity segment, where the supply curve is steeper, shocks to both supply and demand significantly influence quantities. In contrast, inflation shocks explain only a small portion of the variance in quantities in either segment, consistent with both demand and supply shifting upwards simultaneously in response to rises in expected inflation.

Figure 10 ESTIMATED IMPULSE RESPONSE FUNCTIONS TO AN INFLATION SHOCK

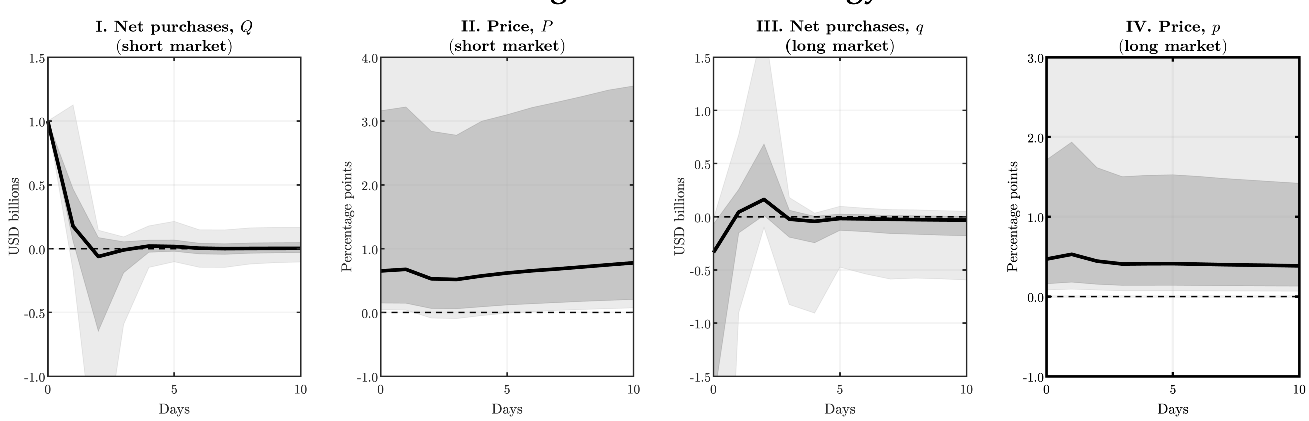
(a) Heteroskedasticity strategy



(b) Granular instrumental variables strategy



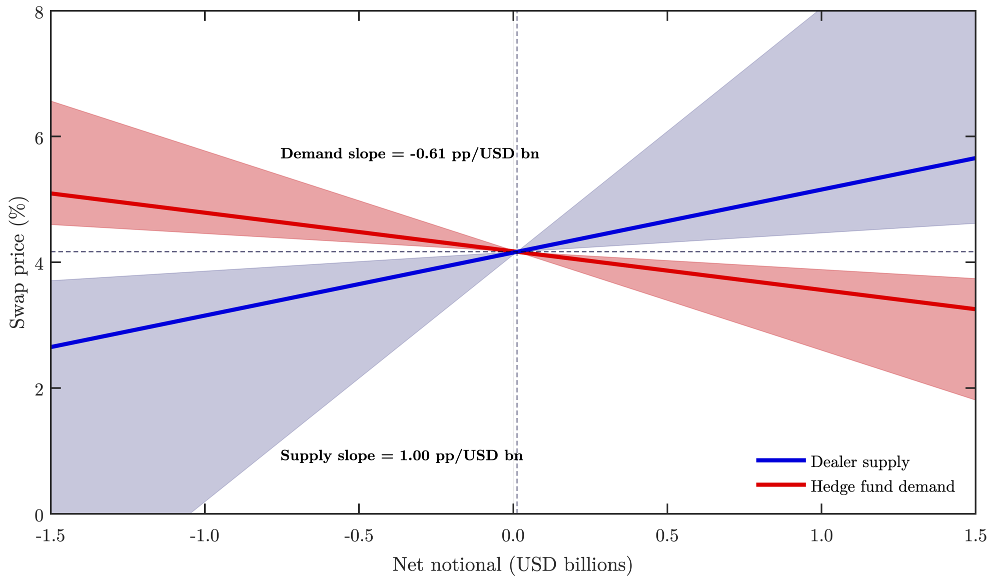
(c) Sign restrictions strategy



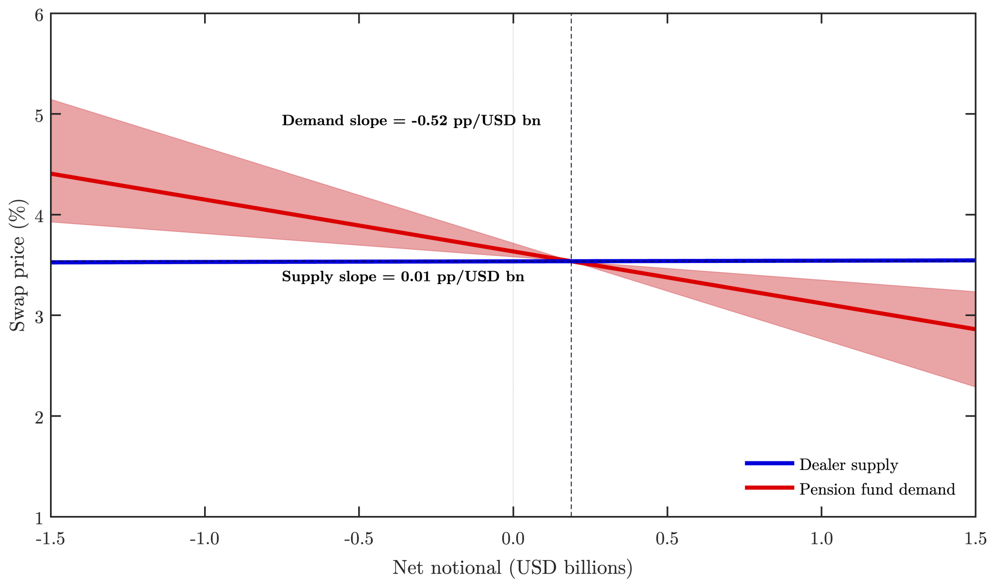
NOTE: Impulse response functions to an inflation shock scaled to raise quantities in the short market, Q, by \$1bn. The three panels present the inflation shock identified from the (*i*) sign restrictions strategy, (*ii*) heteroskedasticity-based strategy and (*iii*) granular identification strategy. In each panel, the bold line indicates the median of the draws from the sampler and shaded bands are its 68% and 90% credible intervals.

Figure 11 ESTIMATED SLOPES OF MARKET DEMAND AND SUPPLY FUNCTIONS

(a) Short maturity UK inflation swap market

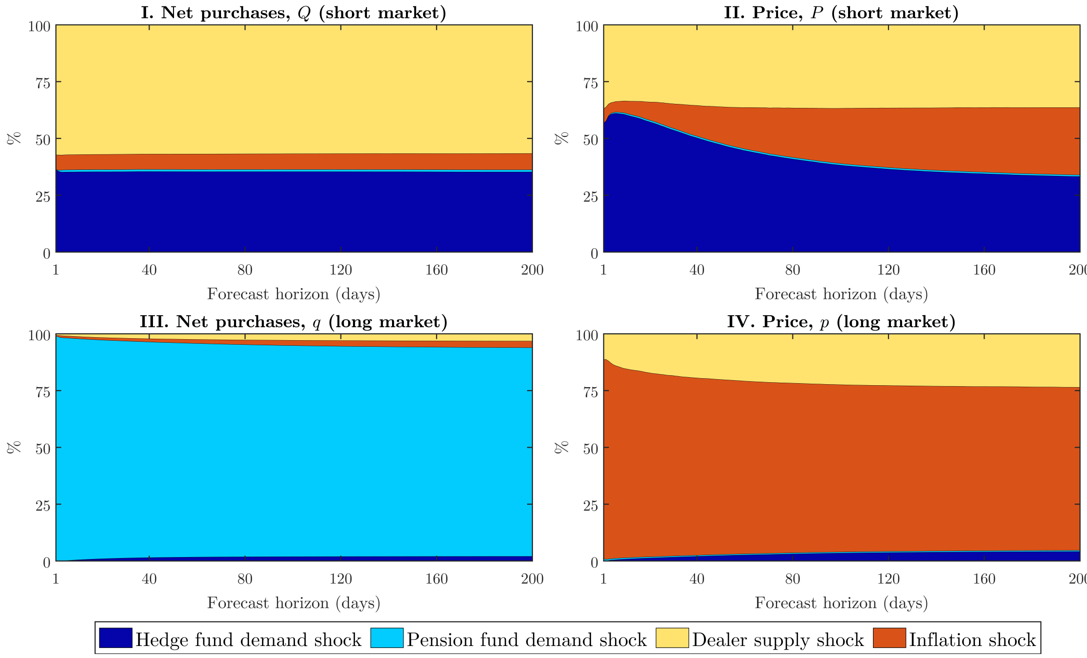


(b) Long maturity UK inflation swap market



NOTE: Panel (a) shows estimates of the supply (blue) and demand (red) curves for the short maturity market segment, defined as $\frac{dP/d\varepsilon_b}{dQ/d\varepsilon_b}$ and $\frac{dP/d\varepsilon_h}{dQ/d\varepsilon_h}$, respectively, using the impulse response functions from the sign restriction strategy. The intercepts of the curves are set so that P and Q are at the sample averages. The panels show the median estimate and 68% credible intervals. Panel (b) repeats the exercise for the long maturity market segment (p and q).

Figure 12 FORECAST ERROR VARIANCE DECOMPOSITION



NOTE: Forecast error variance decompositions for $\mathbf{Y}_t$ based on the VAR model identified using sign restrictions. Median values from the sampler are used.

7 Links to survey expectations and liquidity

This section connects and compares our estimates of expected inflation with those derived from surveys, our estimates of frictional shocks with measures of market liquidity, and examines differences across institutions in terms of disagreement, size, and price impact.

7.1 Institutional trading and beliefs

A longstanding challenge in the literature on inflation expectations is whether survey responses correlate with the actual choices of respondents. By extracting implicit beliefs from trading behavior, we now compare these to surveyed beliefs in a market where substantial sums are at stake.[22]

Constrained by the availability of data, we focus our attention on the short maturity market segment. The sensitivity of a bank's trading to shocks to expected inflation is the coefficient $\beta_{b,i}$ in the regression:

$$\frac{\Delta Q_{b,i,t}}{a_{b,i,t}} = constant_b + \beta_{b,i}\varepsilon_t^{\pi} + v_{b,i,t}\,. \tag{19}$$

According to Proposition 1, in the model this coefficient should equal:

$$\beta_{b,i} = \frac{M_{b,i} - \Lambda}{\gamma_{b,i}\sigma_{\pi}^2(1-\rho_{\pi,d}^2)}\,. \tag{20}$$

That is, conditioning on the same inflation shock, institutions will select different positions, reflecting their disagreement about expected inflation adjusted for risk aversion.

Survey answers about expected inflation at the institution level, $\hat{\Pi}^e_{b,i,t}$, provide a noisy and possibly biased measure of $E_{b,i}(\Pi)$. Since, in the model, $E_{b,i}(\Pi) = M_{b,i}\Pi^e$, the parameter $\phi_{b,i}$ would serve as a proxy for $M_{b,i}$ in the following panel regression:

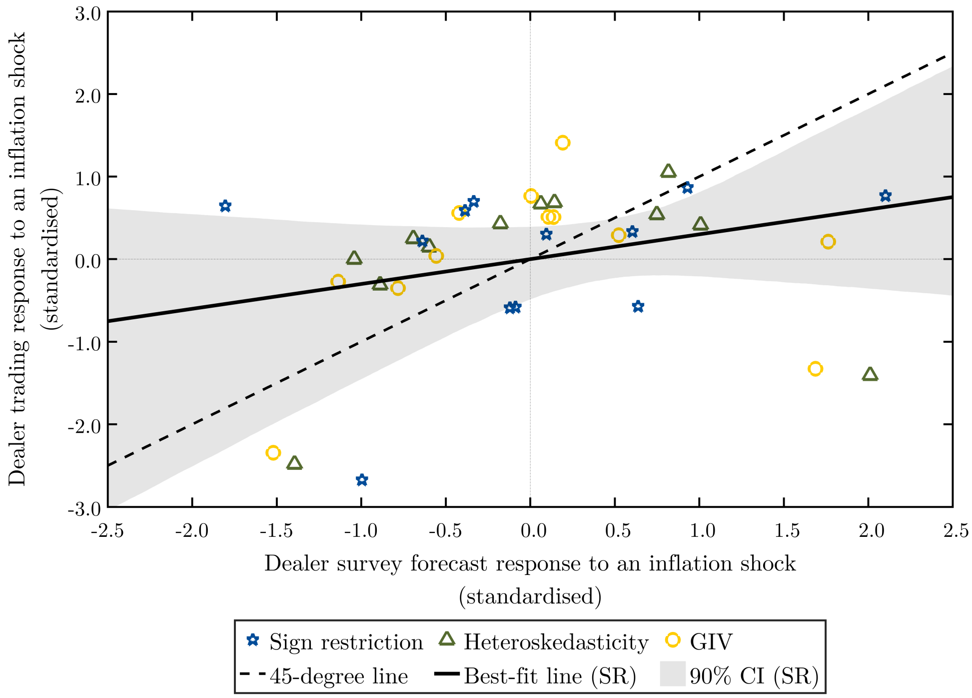
$$\Delta\hat{\Pi}^e_{b,i,t} = constant_b + \phi_{b,i}\varepsilon_t^{\pi} + u_{b,i,t}\,. \tag{21}$$

Hence, under the lenses of our model, the estimated coefficients on the inflation shocks, $\beta_{b,i}$ from equation (19), and $\phi_{b,i}$ from equation (21), should be positively correlated across institutions i, as both are driven by differences in expectations $M_{b,i}$. Beyond the model, a positive association indicates that banks put their money where their mouth is: institutions that revise their inflation expectations upwards following a shock also purchase more inflation protection on the same day.

We implement this test as follows. In both equations (19) and (21), we only have noisy measures of ε_t^{π} identified up to a sign and scale, so we use ΔP_t as the regressor and ε_t^{π} as an instrument. Since the errors $v_{b,i,t}$ are correlated across institutions (because of common frictions, according to the model), and likewise for $u_{b,i,t}$, we use a three-stage least square estimator in a seemingly unrelated regression framework. For equation (19), the left-hand side are changes in net positions

[22] See also Giglio et al. (2021).

Figure 13 EXPECTATIONS AND HETEROGENEITY IN TRADING



NOTE: Scatter plot of estimates of $\beta_{b,i}$ (the sensitivity of trading to expected inflation) on the vertical axis against $\phi_{b,i}$ (the sensitivity of survey answer to expected inflation) on the horizontal axis, standardized to have zero mean and a standard deviation of one. Different colors refer to the three identification strategies. The fitted line refers to the estimates using sign restrictions, with 90% bootstrapped confidence intervals from 10,000 pseudo-samples drawing 12 banks with replacement to re-estimate the two parameters.

(scaled by size) at the institution level, and the sample is the same daily data as in the VAR models. For equation (21), we measure $\hat{\Pi}^{e}_{b,i,t}$ using the monthly responses from a Bloomberg survey of chief economists at dealer banks regarding their expected 1-year ahead RPI inflation. Since the identity of the institution is made public by Bloomberg, we can match it to our (private) data for all but one of the dealer banks in the short market. The sample is now monthly, between January 2019 and February 2023, so we use ΔP_t over the course of the month as the regressor and the monthly cumulative sum of ε^{π}_{t} as the instrument.

Figure 13 shows the scatter plots of the standardized estimates from the two regressions. The coefficients on trading behavior are plotted on the vertical axis, while those on survey answers are on the horizontal axis. The two sets of coefficients are clearly positively and statistically significantly related, despite the small number of matched banks (12). This finding is consistent across different identification strategies. We conclude that surveys provide a meaningful signal of the factors driving actual trading behavior.[23]

[23]Insofar as survey answers reflect subjective beliefs while trading behavior responds to risk-adjusted beliefs (Adam et al., 2021), one interpretation of this result is that disagreement in subjective beliefs across institutions is more quantitatively relevant than differences in risk attitudes or exposures.

7.2 Whose beliefs matter most in the market?

Consider the impact of a change in the inflation expectations of dealer bank i on the price of the short maturity inflation swap, relative to that of another dealer bank i'. Applying Lemma 1 together with the definition of trade-sensitivity in equation (20), this is:

$$RelativePriceImpact_{i,i'} = \frac{\beta_{b,i} a_{b,i}}{\beta_{b,i'} a_{b,i'}}. \tag{22}$$

By multiplying the estimates of $\beta_{b,i}$ that we just discussed by the gross notional positions of the institution $a_{b,i}$, we can assess which institutions are moving the market.

Figure 14 presents the estimates for the four market-segment and institution-type combinations, with different points reflecting the different identification strategies. The top panel shows that, in the short maturity segment of the market, the actions and beliefs of most hedge funds and dealer banks have little impact. However, three to five dealer banks and three to five hedge funds both respond significantly and invest heavily. For some hedge funds, the coefficients are positive, while for others, they are negative. Overall, the average (pooled) coefficient is positive for hedge funds while it is negative for dealers, consistent with assumption B4. Moreover, as Appendix D.4 shows, the correlation between $\beta_{h,i}$ and $a_{h,i}$ is small, and the price impact per institution aligns well with its trade sensitivity. These estimates suggest that institutions disagree in their interpretation of inflation shocks, justifying both the large volume of trade observed and the small net positions aggregated over hedge funds in spite of the large positions of individual institutions.

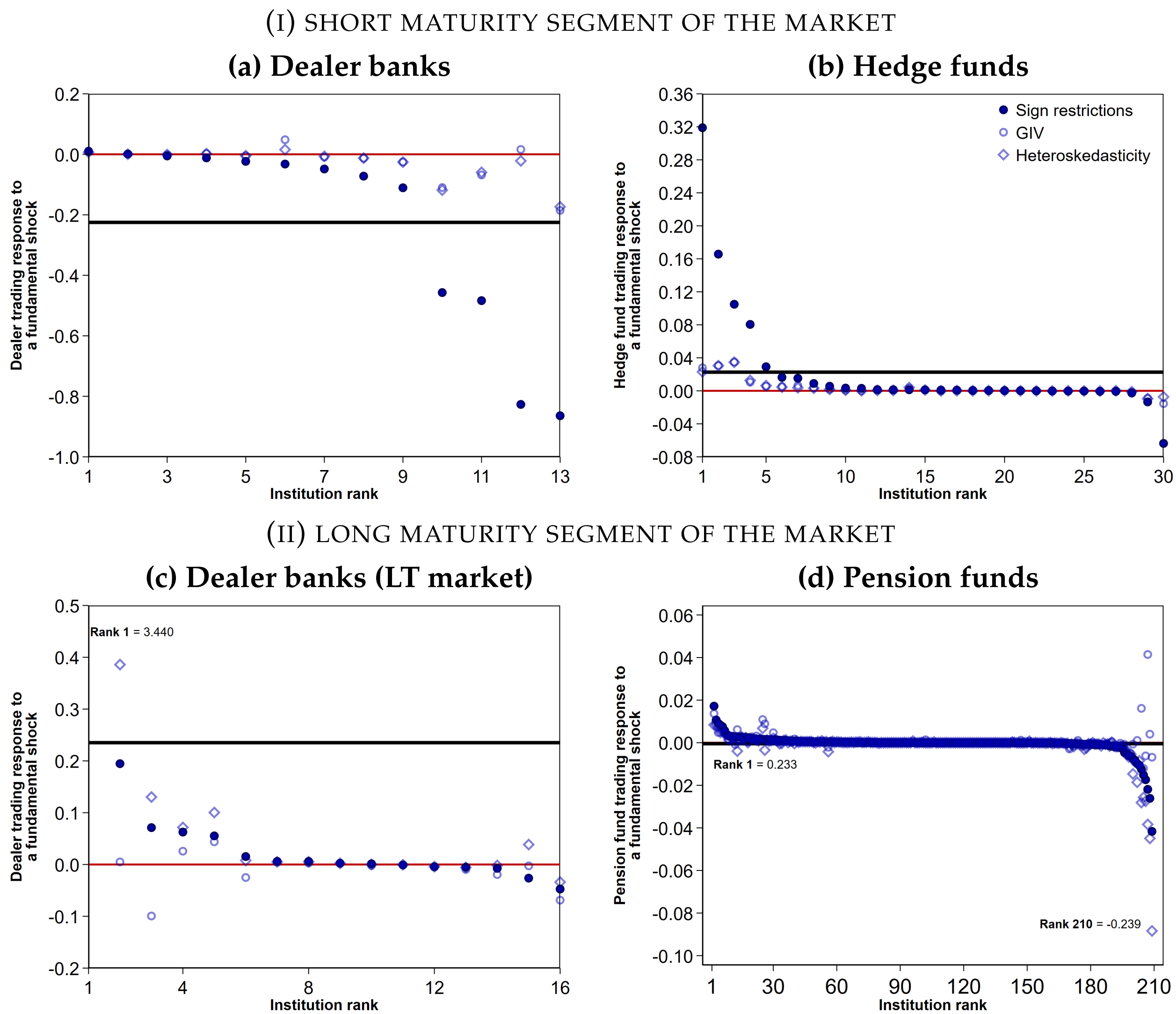
The bottom panel illustrates the price impacts in the long maturity segment of the market. The concentration among dealer banks remains, with the three larger institutions being more responsive (e.g., the estimates of $a_{b,i}$ and $\beta_{b,i}$ are positively correlated), although the relative price impacts are not as pronounced as in the short maturity segment. Among pension funds, the dispersion of price impact is smaller, with a few institutions deviating from the horizontal axis due to their large size (high $a_{f,i}$), as the dispersion of trading impact of expectations $\beta_{f,i}$ is much smaller.

7.3 Frictional shocks and liquidity

Our estimate of the frictions affecting the ability of dealer banks to supply inflation protection is derived from observing trading behavior and prices in markets. We found that shocks to these frictions explain more than half of the variation in quantities in the short maturity segment of the market, and around one-quarter of the variation in prices at both short and long maturities.

Figure 15 compares the component of short maturity inflation swap prices driven by shocks to dealers (ε_b) against a conventional measure of dealer supply constraints: the quoted bid-ask spread in the market. During the significant episodes previously examined, the two measures are significantly positively correlated. Moreover, our estimate of the dealer supply shock typically

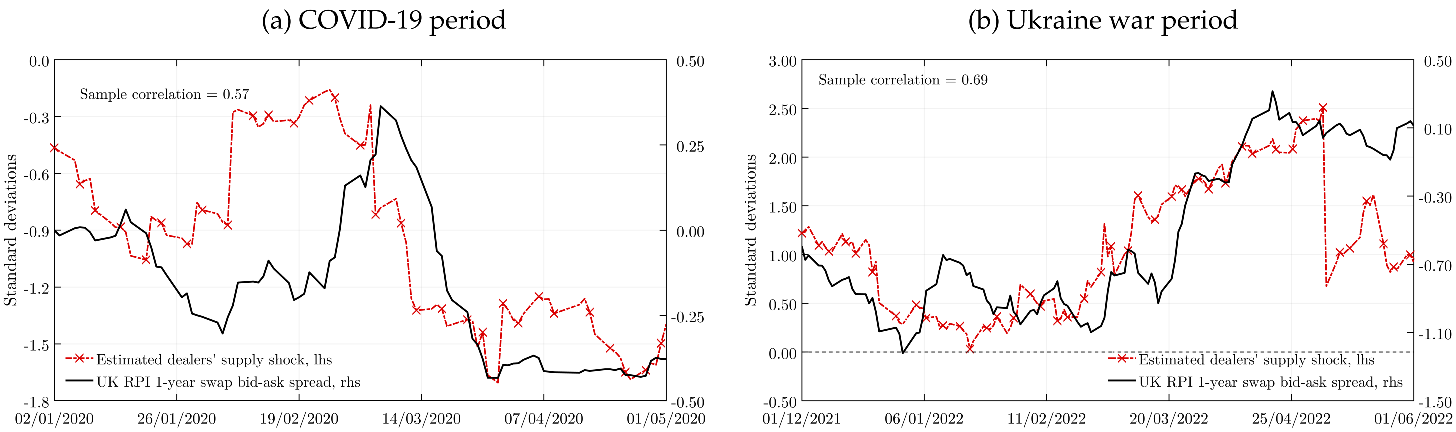
Figure 14 ESTIMATES OF THE PRICE IMPACT OF INSTITUTIONS' INDIVIDUAL EXPECTATIONS



NOTE: Individual markers denote point estimates of β_i across the three different strategies. Institutions are ranked based on the sign restriction strategy. The thick black lines are the average coefficient estimates estimated by three-stage least squares, using our identified inflation shock from the sign restriction strategy as an instrument for the change in the short maturity inflation swap breakeven rates. For panels (c) and (d), the respective responses of the highest and lowest ranking institutions are not plotted and their values are instead indicated on top/bottom of the panels.

moves several days in advance of the bid-ask spread. This both confirms the validity of our estimates and suggests that they may be useful as an alternative measure of restricted supply.

Figure 15 COMPARISON BETWEEN DEALER SUPPLY SHOCKS AND MARKET BID-ASK SPREADS



NOTE: This figure shows the bid-ask spreads quoted on a 1-year zero-coupon UK RPI swap (right scale), and the cumulative contribution to the short horizon inflation swap rate by the dealer supply shock, estimated using the sign restriction identification strategy, taking the median from the sampler (left scale). Both are standardized. Data on UK RPI swap bid-ask spreads are obtained from Bloomberg.

8 Conclusion: summary of findings

This paper complements the public data on inflation swap contract prices, which are extensively used by researchers and policymakers, with valuable new non-public data on the quantities behind these prices and the institutions involved in trades. A model that captures the extreme segmentation observed in this market suggests three new identification strategies to distinguish changes in expected inflation from shocks to the distinct demands of the types of institutions trading in the market. This provides new measures of expected inflation at different horizons for macroeconomists, as well as several features of the functioning of this large market for further study by financial economists. This analysis has led to the following lessons about inflation risk:

First, at short maturities, hedge funds and dealers alternate between negative and positive net positions that average to zero, while at long maturities, dealers consistently provide inflation protection to pension funds. The market for inflation swaps fits into a segmented-markets model, with dealers acting as arbitrageurs.

Second, given the heteroscedasticity in the revelation of inflation news at data release dates, one can utilize the time-series variation in price and quantity data to identify changes in inflation expectations. With data on individual institutions' positions over time, and granularity in the size of those positions, instrumental variables can be constructed for the frictional shocks that shift relative demand and supply in these markets. Using high-frequency data, we can exploit the spillover effects of shocks across segmented markets to identify the entire system of demand and supply. These three alternative approaches provide consistent measures of market participants' inflation expectations at different horizons.

Third, our measures of expected inflation for long horizons (10 years and above) indicate that swap prices have overstated the unanchoring of expectations in our sample. Around significant events, frictions have tended to move in the same direction as expectations, causing prices to overstate changes in inflation expectations. At short maturities, swap prices are unreliable measures of expected inflation, exhibiting large persistent differences between the two. Researchers and policymakers should use swap prices with caution and ideally apply our methods (or others) to filter them and extract accurate signals for monetary policy and macroeconomic analyses.

Fourth, prices in this market appear to fully reflect information within one to three days, and the slope of the supply function for inflation protection at long maturities by dealer banks is nearly horizontal (but not so at short maturities). Consequently, the large fluctuations in quantities traded are almost entirely due to shocks to trading frictions, while expectations account for three-quarters of the movements in prices. At short maturities, frictions affecting hedge funds are nearly as significant as those affecting dealer banks.

Fifth, we observed significant dispersion in beliefs about inflation both within and between types of institutions, and large price impacts from a handful of traders. There is a strong correlation between the inflation expectations of banks, as indicated by their survey responses, and their trading activity in the inflation swap market.

Overall, this paper contributes data for measurement, techniques for estimation, and empirical results that present intriguing 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.

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 11.

ADAM, K., D. MATVEEV, AND S. NAGEL (2021): "Do Survey Expectations of Stock Returns Reflect Risk Adjustments?" *Journal of Monetary Economics*, 117, 723–740.

ANDRADE, P. AND F. FERRONI (2021): "Delphic and Odyssean Monetary Policy Shocks: Evidence from the Euro 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.

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

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.

BEGENAU, J., M. PIAZZESI, AND M. SCHNEIDER (2015): "Banks' Risk Exposures," NBER working paper 21334.

BINDER, C. AND R. KAMDAR (2022): "Expected and Realized Inflation in Historical Perspective," *Journal of Economic Perspectives*, 36, 131–56.

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.

BRETSCHER, L., L. SCHMID, I. SEN, AND V. SHARMA (2021): "Institutional Corporate Bond Pricing," Swiss Finance Institute research paper 21-07.

BRUNNERMEIER, M., D. PALIA, K. A. SASTRY, AND C. A. SIMS (2021): "Feedbacks: Financial Markets and Economic Activity," *American Economic Review*, 111, 1845–79.

CENEDESE, G., P. DELLA CORTE, AND T. WANG (2021): "Currency Mispricing and Dealer Balance Sheets," *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," *Annual Review of Financial Economics*, 15, 433–448.

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 983.

CZECH, R., P. DELLA CORTE, S. HUANG, AND T. WANG (2022): "FX Option Volume," Bank of England Staff working paper 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 (2025): "Getting to the Core: Inflation Risks Within and Across Asset Classes," *Review of Financial Studies*, forthcoming.

FERRARA, G., P. MUELLER, G. VISWANATH-NATRAJ, AND J. WANG (2022): "Central Bank Swap Lines: Micro-level Evidence," Bank of England Staff working paper 977.

FLECKENSTEIN, M., F. A. LONGSTAFF, AND H. LUSTIG (2014): "The TIPS-Treasury Bond Puzzle," *Journal of Finance*, 69, 2151–2197.

GABAIX, X. AND R. S. KOIJEN (2021): "In Search of the Origins of Financial Fluctuations: The Inelastic Markets Hypothesis," NBER working paper 28967.

——— (2024): "Granular Instrumental Variables," *Journal of Political Economy*, 132, 000–000.

GABAIX, X., R. S. KOIJEN, F. MAINARDI, S. S. OH, AND M. YOGO (2025): "Limited Risk Transfer between Investors: A New Benchmark for Macro-Finance Models," NBER working paper 33336.

GIGLIO, S., M. MAGGIORI, J. STROEBEL, AND S. UTKUS (2021): "Five Facts about Beliefs and Portfolios," *American Economic Review*, 111, 1481–1522.

GRANJA, J., E. JIANG, G. MATVOS, T. PISKORSKI, AND A. SERU (2024): "Book Value Risk Management of Banks: Limited Hedging, Htm Accounting, and Rising Interest Rates," NBER working paper 32293.

HANSON, S. G., A. MALKHOZOV, AND G. VENTER (2024): "Demand-and-Supply Imbalance Risk and Long-Term Swap Spreads," *Journal of Financial Economics*, 154, 103814.

HAU, H., P. HOFFMANN, S. LANGFIELD, AND Y. TIMMER (2021): "Discriminatory Pricing of Over-the-Counter Derivatives," *Management Science*, 67, 6660–6677.

JANSEN, K. A. (2025): "Long-term Investors, Demand Shifts, and Yields," *Review of Financial Studies*, 38, 114–157.

JANSEN, K. A., W. LI, AND L. SCHMID (2024): "Granular Treasury Demand with Arbitrageurs," NBER working paper 33243.

KHETAN, U., I. NEAMTU, AND I. SEN (2023): "The Market for Sharing Interest Rate Risk: Quantities Behind Prices," Bank of England Staff working paper 1031.

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 (2024): "Which Investors Matter for Equity Valuations and Expected Returns?" *Review of Economic Studies*, 91, 2387–2424.

KOIJEN, R. S. AND M. YOGO (2020): "Exchange Rates and Asset Prices in a Global Demand System," CEPR discussion paper 14874.

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

LEWIS, D. (2024): "Identification Based on Higher Moments," CeMMAP working paper 03/24.

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 31166.

Nakamura, E. and J. Steinsson (2018): "High-frequency identification of monetary non-neutrality: the information effect," *Quarterly Journal of Economics*, 133, 1283–1330.

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.

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

Appendix – For Online Publication

This appendix presents some further features of the UK inflation swap market data (Section A), proves the lemmas and propositions (Section B), describes the implementation of the granular instrumental variables (Section C), and presents further empirical estimates (Section D). A supplemental appendix contains more supporting information.

A Additional data

This section of the appendix describes features of the swap market that complement Section 2.5.

Section A.1 discusses the connection and differences between inflation swaps and index-linked government bonds. Section A.2 provides more data and information on the gross notional positions in the dealer-client segment of the market to complement the net positions in the main text. Section A.3 shows the equivalent of Figure 4 in the main text for the Euro Area until December 2020 when the UK's exit from the European Union rendered our data unrepresentative. Finally, Section A.4 provides further evidence of market segmentation in the UK RPI market, by showing the volume-weighted median maturity of the executed trades by both hedge funds and pension funds for the most recent period in our data sample.

A.1 Inflation swaps versus linkers

Index-linked government bonds (linkers) and inflation swaps present distinct characteristics that affect their utility in hedging inflation risk.

First, the size and customizability of these instruments differ significantly. Linkers have a limited supply, with approximately $430bnn outstanding in DMO issuance at the end of 2022 (out of which pension funds held around $300bn, see Figure A1). The UK RPI swap market instead has an outstanding notional of around $4tr (Figure A3). Moreover, pension fund liabilities, totaling around $1.2tr at the end of 2022, greatly exceed the outstanding amount of linkers (Figure A1). Inflation swaps offer a synthetic product tailored to suit hedging needs, without being limited by issuance levels. This makes swaps more flexible in terms of matching maturity and potentially more attractive for investors seeking customized inflation protection.

Second, swaps do not require cash funding or upfront payments. In contrast, linkers involve upfront costs associated with purchasing the bonds. Investors with little initial capital or limited liquidity would prefer swaps.

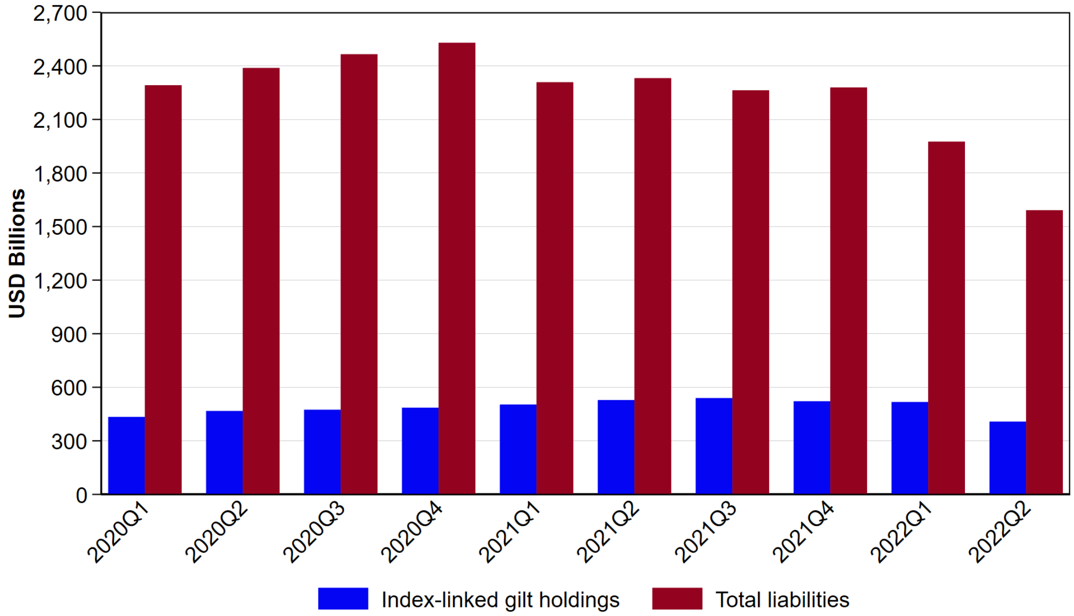
Third, price-insensitive demand for linkers by pension funds results in them being illiquid. This limits their effectiveness for inflation hedging, particularly during periods of market stress. Swaps, instead, can be created easily by dealers, and can be terminated in a straightforward manner by the two parties since their net present value is kept at zero.

Using granular data on gilt transactions from the MiFID II Bond Transaction Database, we calculate the correlation between the monthly change in swap net notionals and the net trading volumes in index-linked gilts for a class of investors. This correlation is small: -0.15 for hedge funds and -0.10 for pension funds. Figure A2 plots the two series. The lack of an association between them suggests that investors may use them for different purposes or at different moments in time.

A.2 Gross positions in the dealer-client inflation swap market

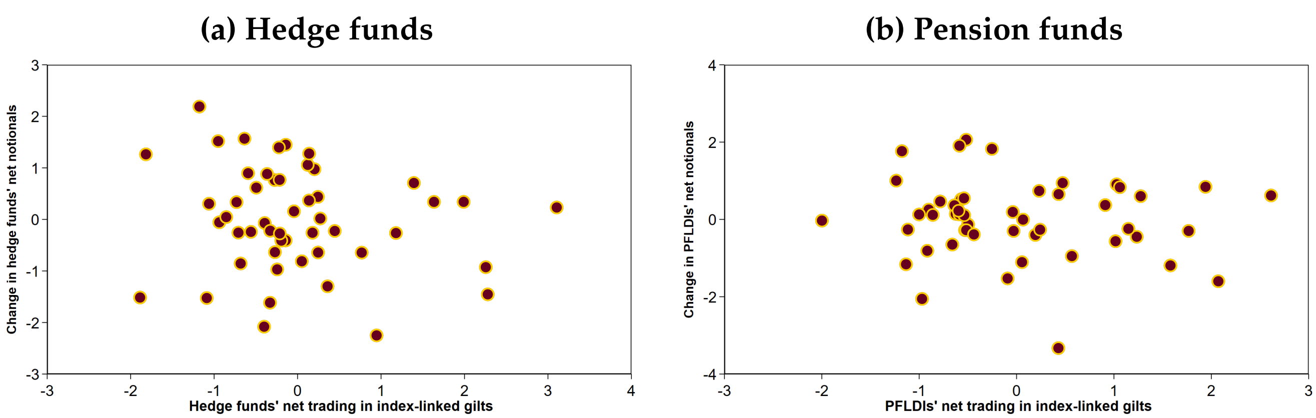
Panel (a) in Figure A3 shows a time series plot of the total gross notional value of all inflation swaps traded in the UK market. As described in the main text, around 22% of these gross notionals are in the dealer client segment that we focus upon. Panel (b) of the figure shows the average market share of different clients of dealer banks based on gross notional positions in the dealer-client segment.

Figure A1 PENSION FUNDS' INDEX-LINKED GILT HOLDINGS VERSUS TOTAL LIABILITIES



NOTE: This figure highlights the persistent gap between the value of index-linked gilts held by defined-benefit pension funds and the value of their total liabilities. A possible explanation for the persistent gap illustrated is the inadequate supply of index-linked gilts in the market, and they are also less attractive relative to inflation swaps as a means to speculate on or hedge against inflation given the upfront capital costs involved. The data spans from 2020Q1 to 2022Q2. SOURCE: Pension Protection Fund PPF 7800 Data and Office for National Statistics.

Figure A2 INFLATION SWAP NET NOTIONALS VS. INDEX-LINKED GILT NET TRADING VOLUME



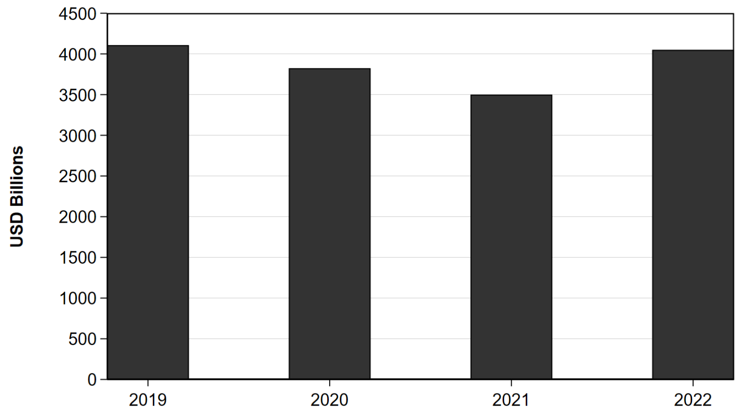
NOTE: The left figure shows the standardised change in monthly hedge fund net notional position in the UK RPI inflation swap market, compared to the standardized monthly hedge fund net trading volumes in the UK index-linked gilt market. The right figure does the same for pension funds. SOURCE: MiFID II Bond Transaction Database & DTCC Trade Repository.

A.3 Net notional positions in the EU inflation market

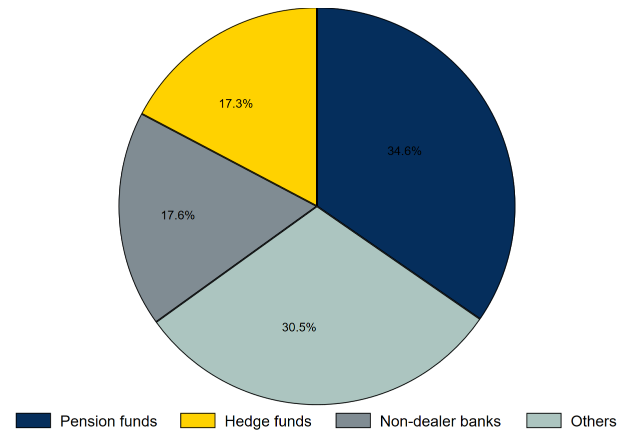
Figure A4 shows equivalent data for Euro Area inflation swaps until December 2020 when the UK's exit from the EU rendered our data unrepresentative. Up until then, the net notional positions traded in the EU inflation swap market (EU HICP index and the EU CPI index) are very similar to the UK RPI swap market, with pension funds being a net receiver of inflation and dealer banks bear inflation risk.

Figure A3 GROSS NOTIONAL VOLUMES: AVERAGE AND COMPOSITIONS

(a) Average volume outstanding

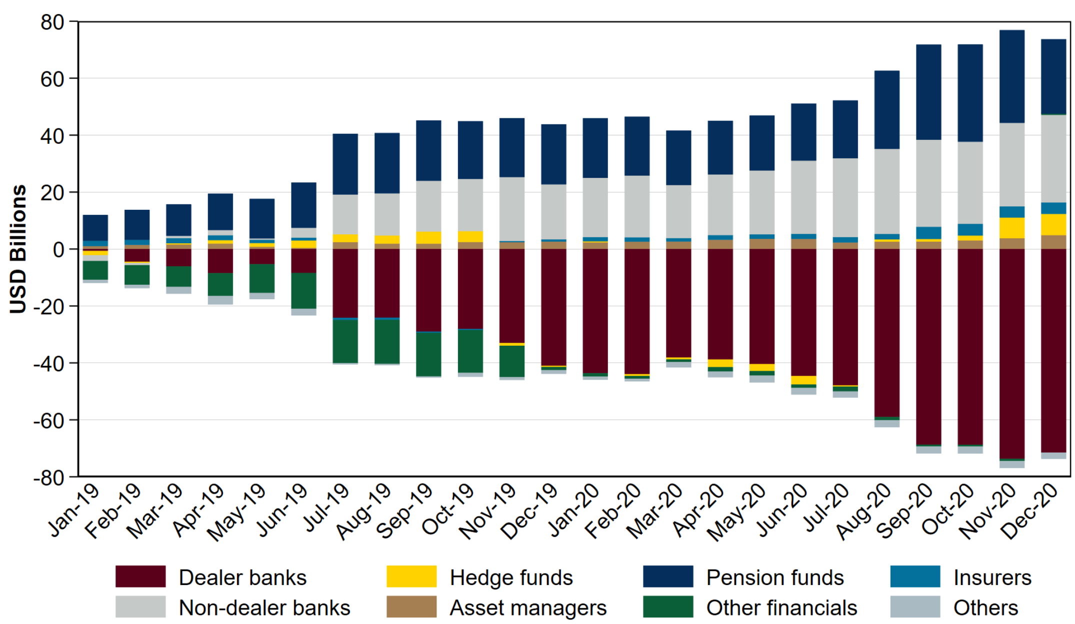


(b) Average market shares



NOTE: Panel (a) 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 month-ends. Panel (b) 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 month-ends in the sample. SOURCE: DTCC Trade Repository.

Figure A4 NET NOTIONAL POSITIONS IN THE EU INFLATION SWAP MARKET



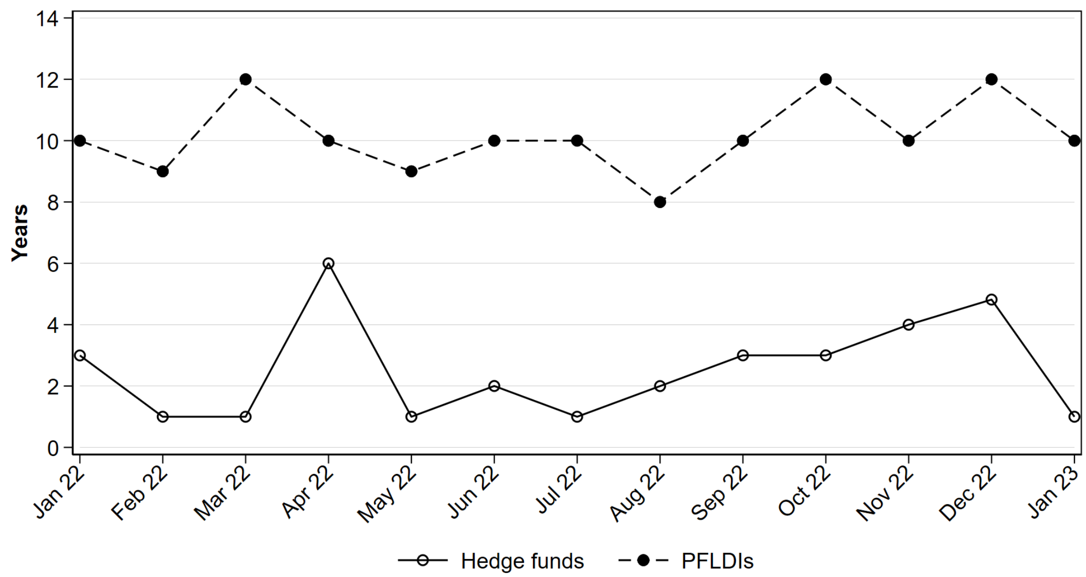
NOTE: The figure shows the aggregate net notional position of all investor-sectors in the dealer-client segment of the EU inflation swap market on a monthly frequency, where each bar measures the outstanding positions at month-end. These positions only include the trading of inflation swaps traded on the EU HICP index and the EU CPI index, and they exclude the positions traded on EU member country-specific inflation indexes. A positive value indicates that the sector is a net buyer of inflation protection. “Others” include: state, supranational, proprietary trading firms, trading services and central banks. The data sample is from January 2019 to December 2020, truncated due to Brexit. SOURCE: DTCC Trade Repository.

A.4 Further evidence of market segmentation

Figure A5 shows the volume-weighted median maturity of the executed trades by both hedge funds and pension funds for the most recent period in our data sample. The median initial time-to-maturity of trades signed by pension funds is approximately 11-years, while that for hedge funds is approximately 2-years.

This complements the evidence shown in the main text that the UK RPI market is segmented, with hedge funds primarily trading inflation risk at short-maturities, while pension funds buy long term inflation protection.

Figure A5 MEDIAN MATURITY OF UK RPI SWAP CONTRACTS TRADED



NOTE: This figure shows the volume-weighted median maturity of the executed trades by both hedge funds and pension funds from January 2019 to February 2023. The median initial time-to-maturity of trades signed by pension funds is approximately 11-years, while that for hedge funds is approximately 2-years, confirming market segmentation across maturities. SOURCE: DTCC Trade Repository.

B Proofs of the model results

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

B.1 Proof of proposition 1

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

$$\boldsymbol{\nu} = \begin{pmatrix} \mathbb{E}_{b,i}[\pi] - p \\ \mathbb{E}_{b,i}[\Pi] - P \\ \mathbb{E}_{b,i}[d] - s \\ \mathbb{E}_{b,i}[y_{b,i}] \end{pmatrix} = \begin{pmatrix} \mu_{b,i}\pi^e - p \\ M_{b,i}\Pi^e - P \\ \theta_d - s \\ 0 \end{pmatrix}, \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_\pi^2 & \sigma_{\pi,\Pi} & \sigma_{\pi,d} & \sigma_{\pi,y_{b,i}} \\ \sigma_{\Pi,\pi} & \sigma_\Pi^2 & \sigma_{\Pi,d} & \sigma_{\Pi,y_{b,i}} \\ \sigma_{d,\pi} & \sigma_{d,\Pi} & \sigma_d^2 & 0 \\ \sigma_{y_{b,i},\pi} & \sigma_{y_{b,i},\Pi} & 0 & \sigma_{y_{b,i}}^2 \end{pmatrix}. \tag{A1}$$

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

$$\mathbb{E}_{b,i}[\Delta a_{b,i}] = (\mu_{b,i}\pi^e - p)q_{b,i} + (M_{b,i}\Pi^e - P)Q_{b,i} + (\theta_d - s)e_{b,i}, \tag{A2}$$

$$\begin{aligned} Var_{b,i}[\Delta a_{b,i}] = {} & q_{b,i}^2\sigma_\pi^2 + Q_{b,i}^2\sigma_\Pi^2 + e_{b,i}^2\sigma_d^2 + \sigma_{y_{b,i}}^2 \\ & + 2q_{b,i}Q_{b,i}\sigma_{\pi,\Pi} + 2q_{b,i}e_{b,i}\sigma_{\pi,d} + 2q_{b,i}\sigma_{\pi,y_{b,i}} + 2Q_{b,i}e_{b,i}\sigma_{\Pi,d} + 2Q_{b,i}\sigma_{\Pi,y_{b,i}}. \end{aligned} \tag{A3}$$

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

$$\mathbb{E}_{b,i}\Big[-exp\big(-\widetilde{\gamma}_{b,i}a'_{b,i}\big)\Big] = -exp\left\{-\widetilde{\gamma}_{b,i}\left[a_{b,i} + \mathbb{E}_{f,i}[\Delta a_{b,i}] - \frac{\widetilde{\gamma}_{b,i}Var_{b,i}[\Delta a_{b,i}]}{2}\right]\right\}. \tag{A4}$$

The bank's optimization is then equivalent to a mean-variance problem:

$$\begin{aligned}\max_{\{q_{b,i},Q_{b,i},e_{b,i}\}} \quad & \Big[(\mu_{b,i}\pi^e - p)q_{b,i} + (M_{b,i}\Pi^e - P)Q_{b,i} + (\theta_d - s)e_{b,i}\Big] - \frac{\widetilde{\gamma}_{b,i}}{2}\Big[q^2_{b,i}\sigma^2_\pi + Q^2_{b,i}\sigma^2_\Pi + e^2_{b,i}\sigma^2_d \\ & + \sigma^2_{y_{b,i}} + 2q_{b,i}Q_{b,i}\sigma_{\pi,\Pi} + 2q_{b,i}e_{b,i}\sigma_{\pi,d} + 2q_{b,i}\sigma_{\pi,y_{b,i}} + 2Q_{b,i}e_{b,i}\sigma_{\Pi,d} + 2Q_{b,i}\sigma_{\Pi,y_{b,i}}\Big], \\ & \text{subject to:} \quad G^S_b(Q_{b,i}, q_{b,i}, z_{b,i}) \geq 0 \quad \text{and} \quad G^L_b(q_{b,i}, Q_{b,i}, z_{b,i}) \geq 0. \end{aligned} \tag{A5}$$

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

$$q^*_{b,i} = \frac{\mu_{b,i}\pi^e - p^*}{\widetilde{\gamma}_{b,i}\sigma^2_\pi} - \frac{\sigma_{\pi,\Pi}}{\sigma^2_\pi}Q^*_{b,i} - \frac{\sigma_{\pi,d}}{\sigma^2_\pi}e^*_{b,i} - \frac{\sigma_{\pi,y_{b,i}}}{\sigma^2_\pi} - \frac{\lambda^{*L}_{b,i}g^{*L}_{b,i}}{\widetilde{\gamma}_{b,i}\sigma^2_\pi}, \tag{A6}$$

$$Q^*_{b,i} = \frac{M_{b,i}\Pi^e - P^*}{\widetilde{\gamma}_{b,i}\sigma^2_\Pi} - \frac{\sigma_{\pi,\Pi}}{\sigma^2_\Pi}q^*_{b,i} - \frac{\sigma_{\Pi,d}}{\sigma^2_\Pi}e^*_{b,i} - \frac{\sigma_{\Pi,y_{b,i}}}{\sigma^2_\Pi} - \frac{\lambda^{*S}_{b,i}g^{*S}_{b,i}}{\widetilde{\gamma}_{b,i}\sigma^2_\Pi}, \tag{A7}$$

$$e^*_{b,i} = \frac{\theta_d - s}{\widetilde{\gamma}_{b,i}\sigma^2_d} - \frac{\sigma_{\pi,d}}{\sigma^2_d}q^*_{b,i} - \frac{\sigma_{\Pi,d}}{\sigma^2_d}Q^*_{b,i}. \tag{A8}$$

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:

$$q^*_{b,i} = \frac{\mu_{b,i}\pi^e - p^*}{\widetilde{\gamma}_{b,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} - \frac{\sigma_d}{\sigma_\pi}\left[\frac{\theta_d - s}{\widetilde{\gamma}_{b,i}\sigma^2_d(1-\rho^2_{\pi,d})}\right]\rho_{\pi,d} - \frac{\sigma_\Pi}{\sigma_\pi}\left[\frac{\rho_{\pi,\Pi} - \rho_{\pi,d}\rho_{\Pi,d}}{1-\rho^2_{\pi,d}}\right]Q^*_{b,i} \tag{A9}$$

$$- \left[\frac{1}{(1-\rho^2_{\pi,d})\sigma^2_\pi}\right]\left(\sigma_{\pi,y_{b,i}} + \frac{\lambda^{*L}_{b,i}g^{*L}_{b}}{\widetilde{\gamma}_{b,i}}\right), \tag{A10}$$

$$Q^*_{b,i} = \frac{M_{b,i}\Pi^e - P^*}{\widetilde{\gamma}_{b,i}\sigma^2_\Pi(1-\rho^2_{\Pi,d})} - \frac{\sigma_d}{\sigma_\Pi}\left[\frac{\theta_d - s}{\widetilde{\gamma}_{b,i}\sigma^2_d(1-\rho^2_{\Pi,d})}\right]\rho_{\Pi,d} - \frac{\sigma_\pi}{\sigma_\Pi}\left[\frac{\rho_{\Pi,\pi} - \rho_{\Pi,d}\rho_{\pi,d}}{1-\rho^2_{\Pi,d}}\right]q^*_{b,i} \tag{A11}$$

$$- \left[\frac{1}{(1-\rho^2_{\Pi,d})\sigma^2_\Pi}\right]\left(\sigma_{\Pi,y_{b,i}} + \frac{\lambda^{*S}_{b,i}g^{*S}_{b,i}}{\widetilde{\gamma}_{b,i}}\right). \tag{A12}$$

From the expressions above, it becomes clear that if $\rho_{\pi,\Pi} = \rho_{\pi,d}\rho_{\Pi,d}$ as we assumed, then $Q^*_{b,i}$ (and its exogenous shifters) are not part of the solution for $q^*_{b,i}$ and vice versa. Using the definition for $\widetilde{\gamma}_{b,i} = \gamma_{b,i}/a_{b,i}$ we get

$$\frac{q^*_{b,i}}{a_{b,i}} = \frac{\mu_{b,i}\pi^e - p^*}{\gamma_{b,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} - \frac{\sigma_d}{\sigma_\pi}\left[\frac{\theta_d - s}{\gamma_{b,i}\sigma^2_d(1-\rho^2_{\pi,d})}\right]\rho_{\pi,d} - \left[\frac{1}{(1-\rho^2_{\pi,d})\sigma^2_\pi}\right]\left(\frac{\sigma_{\pi,y_{b,i}}}{a_{b,i}} + \frac{\lambda^{*L}_{b,i}g^{*L}_{b,i}}{\gamma_{b,i}}\right), \tag{A13}$$

$$\frac{Q^*_{b,i}}{a_{b,i}} = \frac{M_{b,i}\Pi^e - P^*}{\gamma_{b,i}\sigma^2_\Pi(1-\rho^2_{\Pi,d})} - \frac{\sigma_d}{\sigma_\Pi}\left[\frac{\theta_d - s}{\gamma_{b,i}\sigma^2_d(1-\rho^2_{\Pi,d})}\right]\rho_{\Pi,d} - \left[\frac{1}{(1-\rho^2_{\Pi,d})\sigma^2_\Pi}\right]\left(\frac{\sigma_{\Pi,y_{b,i}}}{a_{b,i}} + \frac{\lambda^{*S}_{b,i}g^{*S}_{b,i}}{\gamma_{b,i}}\right). \tag{A14}$$

This completes the proof of the proposition.

B.2 Pension and hedge fund demand

Given that pension funds also have CARA utility, following the same steps as in Appendix B.1, they solve a mean-variance maximization problem:

$$\max_{\{q_{f,i},e_{f,i}\}} \ (\mu_{f,i}\pi^e - p)q_{f,i} + (\theta_d - s)e_{f,i} - \frac{\widetilde{\gamma}_{f,i}}{2}\left[q_{f,i}^2\sigma_\pi^2 + e_{f,i}^2\sigma_d^2 + \sigma_{y_{f,i}}^2 + 2q_{f,i}e_{f,i}\sigma_{\pi,d} + 2q_{f,i}\sigma_{\pi,y_{f,i}}\right],$$
$$s.t. \ G_f(q_{f,i}, z_{f,i}) \geq 0. \tag{A15}$$

The first-order conditions necessary to attain a maximum are given by:

$$q_{f,i}^* = \frac{\mu_{f,i}\pi^e - p^*}{\widetilde{\gamma}_{f,i}\sigma_\pi^2} - \frac{\sigma_{\pi,d}}{\sigma_\pi^2}e_{f,i}^* - \frac{\sigma_{\pi,y_{f,i}}}{\sigma_\pi^2} - \frac{\lambda_{f,i}^* G_f'(q_{f,i}^*, z_{f,i})}{\widetilde{\gamma}_{f,i}\sigma_\pi^2}, \tag{A16}$$

$$e_{f,i}^* = \frac{\theta_d - s}{\widetilde{\gamma}_{f,i}\sigma_d^2} - \frac{\sigma_{\pi,d}}{\sigma_d^2}q_{f,i}^*, \tag{A17}$$

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

$$\lambda_{f,i}^* G_f(q_{f,i}^*, z_{f,i}) = 0 \quad \text{and} \quad G_f(q_{f,i}^*, z_{f,i}) \geq 0 \quad \text{and} \quad \lambda_{f,i}^* \geq 0, \tag{A18}$$

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

$$\frac{q_{f,i}^*}{a_{f,i}} = \frac{\mu_{f,i}\pi^e - p^*}{\gamma_{f,i}\sigma_\pi^2(1-\rho_{\pi,d}^2)} - \frac{\sigma_d}{\sigma_\pi}\left[\frac{\theta_d - s}{\gamma_{f,i}\sigma_d^2(1-\rho_{\pi,d}^2)}\right]\rho_{\pi,d} - \left[\frac{1}{(1-\rho_{\pi,d}^2)\sigma_\pi^2}\right]\left[\frac{\sigma_{\pi,y_{f,i}}}{a_{f,i}} + \frac{\lambda_{f,i}^* g_f(q_{f,i}^*, z_{f,i})}{\gamma_{f,i}}\right], \tag{A19}$$

where $\rho_{\pi,d} = \sigma_{\pi,d}/\sigma_\pi\sigma_d$ is the correlation between π and d.

Following precisely the same steps, and for completeness, the demand for short maturity swaps by hedge funds is:

$$\frac{Q_{h,i}^*}{a_{h,i}} = \frac{M_{h,i}\Pi^e - p^*}{\gamma_{h,i}\sigma_\Pi^2(1-\rho_{\Pi,d}^2)} - \frac{\sigma_d}{\sigma_\Pi}\left[\frac{\theta_d - s}{\gamma_{h,i}\sigma_d^2(1-\rho_{\Pi,d}^2)}\right]\rho_{\Pi,d} - \left[\frac{1}{(1-\rho_{\Pi,d}^2)\sigma_\Pi^2}\right]\left[\frac{\sigma_{\Pi,y_{h,i}}}{a_{h,i}} + \frac{\lambda_{h,i}^* g_h(Q_{h,i}^*, z_{h,i})}{\gamma_{h,i}}\right]. \tag{A20}$$

B.3 Proofs of lemma 1 and lemma 2

The proof of lemma 2 follows directly from replacing the asset demand functions in proposition B.2 (or Appendix B.1) and in Appendix B.2 into the market clearing conditions in (7) and rearranging. The proof of 1 comes from setting the frictional components to zero: $\sigma_{\pi,y_{b,i}} = \sigma_{\pi,y_{f,i}} = \sigma_{\pi,y_{h,i}} = \lambda_{b,i}^{L,*} = \lambda_{f,i}^* = 0$.

For completeness, the equilibrium frictionless price in the short maturity segment of the market is:

$$\widetilde{P} = \left[\frac{\sum_{i\in\Theta_h}\widetilde{\gamma}_{h,i}^{-1}M_{h,i} + \sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}M_{b,i}}{\sum_{i\in\Theta_h}\widetilde{\gamma}_{h,i}^{-1} + \sum_{i\in\Theta_b}\widetilde{\gamma}_{b,i}^{-1}}\right]\Pi^e - \frac{\theta_d - s}{\sigma_d\sigma_\Pi}\rho_{\Pi,d}. \tag{A21}$$

C Identification with granular instrumental variables

We provide more details on the identification strategy using granular instrumental variables here.

To start, Section C.1 derives the decomposition of each institution's demand into an idiosyncratic component and a market-wide factors that are to be estimated. This corresponds to equations (11) presented in the main text. Section C.3 explains why the instrument is valid in theory and Section C.4 describes how the three instruments identify the complete system. Section C.2 first looks at the data to find support for granularity in the data. Section C.5 explains how we estimate the factor model using Bai (2009)'s interactive fixed-effects model.

C.1 Terms in the factor model in equation (11)

For exposition, consider the case of identifying the demand shock in the long maturity inflation swap market. We have to find a granular IV for ε_f.

Recall that demand from institution i among pension funds f, and append to it a time index t:

$$\frac{q^*_{f,i,t}}{a_{f,i,t}} = \frac{\mu_{f,i}\pi^e_t - p^*_t}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} - \left(\frac{\sigma_d}{\sigma_\pi}\right)\left[\frac{\theta_{d,t}-s_t}{\gamma_{f,i}\sigma^2_d(1-\rho^2_{\pi,d})}\right]\rho_{\pi,d} \underbrace{-\left[\frac{1}{(1-\rho^2_{\pi,d})\sigma^2_\pi}\right]\left(\frac{\sigma_{\pi,y_{f,i,t}}}{a_{f,i,t}} + \frac{\lambda^*_{f,i,t}g^*_{f,i,t}}{\gamma_{f,i}}\right)}_{=\varepsilon_{f,i,t}}. \tag{A22}$$

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 and frictions:

$$p^*_t = \Lambda\pi^e_t - rp^*_t + lp^*_t\,. \tag{A23}$$

The components are defined as:

$$\Lambda = \frac{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t}\mu_{f,i}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t}+\sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} + \frac{\sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}\mu_{b,i}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t}+\sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}}, \tag{A24}$$

$$rp_t = \left(\frac{\theta_{d,t}-s_t}{\sigma^2_d}\right)\sigma_{\pi,d}\,, \tag{A25}$$

$$lp^*_t = -\frac{\sum_{i\in\Theta_b}\left(\sigma_{\pi,y_{b,i}} + \frac{\lambda^{*L}_{b,i,t}g^{*L}_{b,i,t}}{\widetilde{\gamma}_{b,i,t}}\right)}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t}+\sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} - \frac{\sum_{i\in\Theta_f}\left(\sigma_{\pi,y_{f,i}} + \frac{\lambda^*_{f,i,t}g^*_{f,i,t}}{\widetilde{\gamma}_{f,i,t}}\right)}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t}+\sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}}. \tag{A26}$$

Then, define $\tilde{\varepsilon}_{f,i,t}$ as the idiosyncratic component of the fund-specific demand shock $\varepsilon_{f,i,t}$:

$$\tilde{\varepsilon}_{f,i,t} = \varepsilon_{f,i,t} - \frac{\kappa^{lp}_{f,i}lp^*_t}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})}, \tag{A27}$$

where $\kappa^{lp}_{f,i}$ captures the contribution of pension fund i's liquidity frictions to the market-wide frictions lp^*_t that was defined in equation (A26).

Substituting (A23), (A25) and (A27) into equation (A22) and rearranging terms, the demand system above can be rewritten as:

$$\frac{q^*_{f,i,t}}{a_{f,i,t}} = \frac{(\mu_{f,i}-\Lambda)\pi^e_t + (\kappa^{lp}_{f,i}-1)lp^*_t}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})} + \widetilde{\varepsilon}_{f,i,t} = \boldsymbol{\omega}'_{f,i}\mathbf{F}_t + \widetilde{\varepsilon}_{f,i,t}, \tag{A28}$$

The $\mathbf{F}_t$ are unobserved common factors and the $\boldsymbol{\omega}_{f,i}$ are the fund-specific factor loadings, defined by (where Λ is as defined in equation (8)):

$$\mathbf{F}_t = \begin{pmatrix}\pi^e_t\\ lp^*_t\end{pmatrix}, \quad \boldsymbol{\omega}_{f,i} = \begin{pmatrix}\frac{(\mu_{f,i}-\Lambda)}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})}\\ \frac{\kappa^{lp}_{f,i}-1}{\gamma_{f,i}\sigma^2_\pi(1-\rho^2_{\pi,d})}\end{pmatrix}. \tag{A29}$$

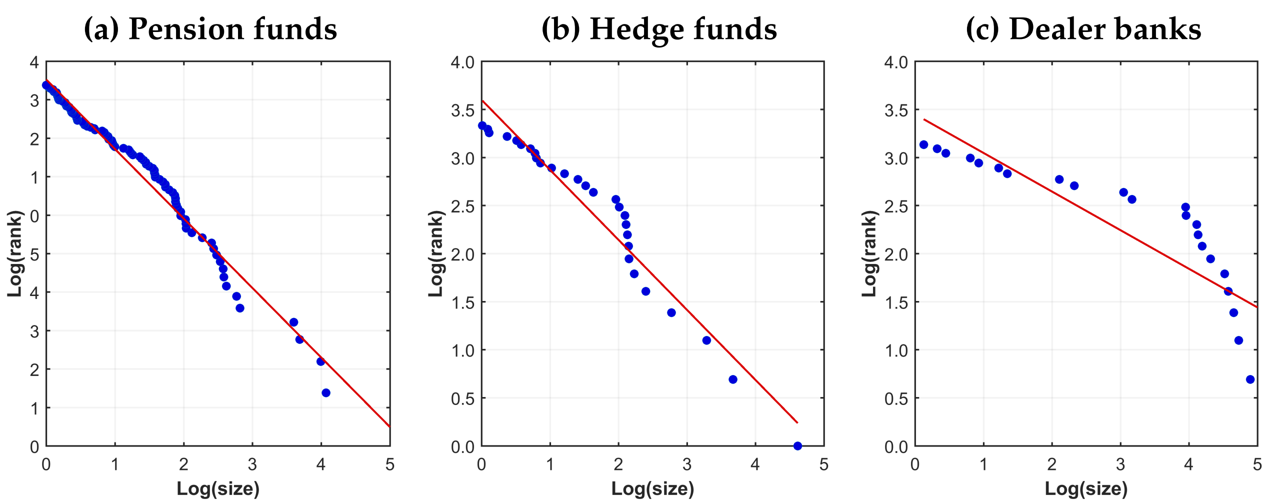
Note that the risk premium term, rp_t, cancels in the fund level demand equation so does not end up being a factor. Note also that the impact of $\varepsilon_{h,t}$ on the long market is spanned by lp_t^*, and likewise for the impact of $\varepsilon_{f,t}$ on the short market.

C.2 Power laws for trading volumes

In our sample, there are 210 pension funds (including liability-driven investment funds). Panel (a) of Figure A6 shows the plot of the rank of each pension fund against their outstanding gross notional positions. For pension funds 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. 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 pension funds' 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.

We repeat this exercise for hedge funds and banks in panels (b) and (c). 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.

Figure A6 INSTITUTIONAL RANK VERSUS OUTSTANDING GROSS NOTIONAL POSITIONS



NOTE: Size refers to the gross notional position outstanding that the institution has acquired in the UK RPI market up to the latest date of our data sample. We estimate it 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. There are a total of 210 pension funds, 30 hedge funds and 16 dealer banks in our data sample. Each scatter marker refers to a given institution, and the line in red denotes the fitted value. SOURCE: DTCC Trade Repository.

C.3 Instrument validity

The granular IV for pension fund demand is given by:

$$GIV_{f,t} = \sum_{i \in \Theta_f} a_{f,i,t} \tilde{\varepsilon}_{f,i,t}. \tag{A30}$$

Recall from equation (9) that the impact of the sector-wide demand frictions from pension funds on the swap price is given by:

$$-\frac{\sum_{i \in \Theta_f} \left\{ \sigma_{\pi, y_{f,i}} + \frac{\lambda^*_{f,i} g^*_{f,i,t}}{\tilde{\gamma}_{f,i,t}} \right\}}{\sum_{i \in \Theta_f} \tilde{\gamma}^{-1}_{f,i,t} + \sum_{i \in \Theta_b} \tilde{\gamma}^{-1}_{b,i,t}}. \tag{A31}$$

Equation (A27) can be rewritten as:

$$\frac{\sum_{i\in\Theta_f}(\kappa^{lp}_{f,i} lp^*_t)\widetilde{\gamma}^{-1}_{f,i,t}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t}+\sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}} + (1-\rho^2_{\pi,d})\sigma^2_\pi \frac{GIV_{f,t}}{\sum_{i\in\Theta_f}\widetilde{\gamma}^{-1}_{f,i,t}+\sum_{i\in\Theta_b}\widetilde{\gamma}^{-1}_{b,i,t}}. \tag{A32}$$

These formulas make it clear that $GIV_{f,t}$ qualifies as a relevant instrumental variable for $\varepsilon_{f,t}$ as long as there is some granularity in $a_{f,i,t}$ so that $GIV_{f,t} \neq 0$. No granularity means $GIV_{f,t} = 0$ as all the idiosyncratic shocks average out.

In terms of satisfying the exclusion restriction, $\tilde{\varepsilon}_{fi,t} \perp \varepsilon_{h,t}, \varepsilon_{b,t}, \varepsilon_{\pi,t}$ by construction, since these three shocks are spanned by $F_{f,t}$. Hence, $\mathbb{E}(GIV_{f,t}\varepsilon_{h,t}) = 0$, $\mathbb{E}(GIV_{f,t}\varepsilon_{b,t}) = 0$ and $\mathbb{E}(GIV_{f,t}\varepsilon_{\pi,t}) = 0$.

The same applies to hedge funds and dealer banks to obtain $GIV_{h,t}$ and $GIV_{b,t}$ as instruments for $\varepsilon_{h,t}$ and $\varepsilon_{b,t}$. These are the demand shocks in the short maturity market and shocks to dealers' supply functions.

C.4 Recovering the inflation shock

Recall that the static representation of the system is:

$$\begin{pmatrix} Q_t \\ P_t \\ q_t \\ p_t \end{pmatrix} = constant + \underbrace{\begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \mathbf{b}_3 & \mathbf{b}_4 \end{bmatrix} \begin{pmatrix} \varepsilon_{h,t} \\ \varepsilon_{f,t} \\ \varepsilon_{b,t} \\ \varepsilon_{\pi,t} \end{pmatrix}}_{\mathbf{B}} = constant + \begin{pmatrix} u_{1,t} \\ u_{2,t} \\ u_{3,t} \\ u_{4,t} \end{pmatrix}, \tag{A33}$$

where each $\mathbf{b}_i$ for $i \in \{1,2,3,4\}$ is a 4×1 column vector. We use $GIV_{f,t}$, $GIV_{h,t}$ 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 (call it $IV_{\pi,t}$) is then a valid instrument for $\varepsilon_{\pi,t}$. To see this, note that the matrix B by assumption can be inverted to obtain:

$$u_{4,t} = a_{4,1}u_{1,t} + a_{4,2}u_{2,t} + a_{4,3}u_{3,t} + \varepsilon_{\pi,t}. \tag{A34}$$

The residuals from the instrumented regression exactly yield $\varepsilon_{\pi,t}$.[24]

Projecting u_t sequentially on $GIV_{f,t}$, $GIV_{h,t}$, $GIV_{b,t}$ and $IV_{\pi,t}$ identifies the coefficients of the structural impact matrix b_1 to b_4 up to sign and scale.

C.5 Estimation of the factor model

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

$$\frac{q_{f,i,t}}{a_{f,i,t}} = \alpha_{f,i} + \tau_t + \sum_{j=1}^{J}\beta_j \frac{q_{f,i,t-j}}{a_{f,i,t-j}} + \boldsymbol{\omega}'_{f,i}\mathbf{F}_t + \widetilde{\varepsilon}_{f,i,t}. \tag{A35}$$

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

Although our model implies a two-factor structure, 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., private versus public or defined contribution versus defined benefit pension funds) or specialization across dealers. We take a two-step approach. First, we select an initial number of factors according to the factor selection criteria in Bai and Ng (2002). We then incrementally add an additional factor until the structural responses of the

[24] 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.

demand and supply shocks satisfy the typical responses associated with these shocks in the specific market of interest. We impose no ex-ante restriction on the size or direction of spillovers across markets or on how the inflation shock we eventually identify affects the system. In this way, our factor selection criteria does not rely on assumptions B3 or B4. This procedure led to 21 factors being estimated for pension fund demand, 12 for hedge fund demand and 9 for dealer bank supply.

D Additional empirical results

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

D.1 Demand and supply slopes using GIV strategy

Figure A7 shows the slopes of demand and supply graphically, together with their credible sets. The relative flatness of the supply curve in the long market versus the short market is also confirmed using this strategy. The estimated demand curves are flatter. However, as discussed, this strategy struggles to precisely estimate the impact of a supply shock, which means that the credible sets over the demand curve slopes are wide.

D.2 Movements around large shocks

Figure A8 shows the actual swap prices around the two large events discussed in Section 5

The supplemental Appendix decomposes the frictional component into its three parts, and shows the decomposition of traded quantities as well.

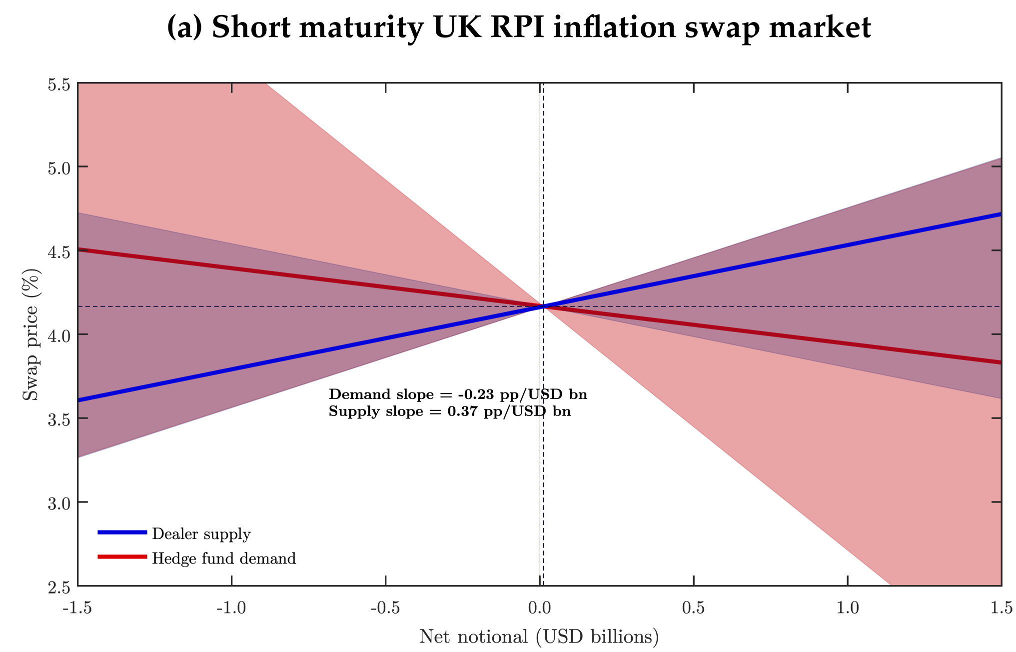
D.3 Autumn 2022 energy price guarantees and the UK LDI crisis

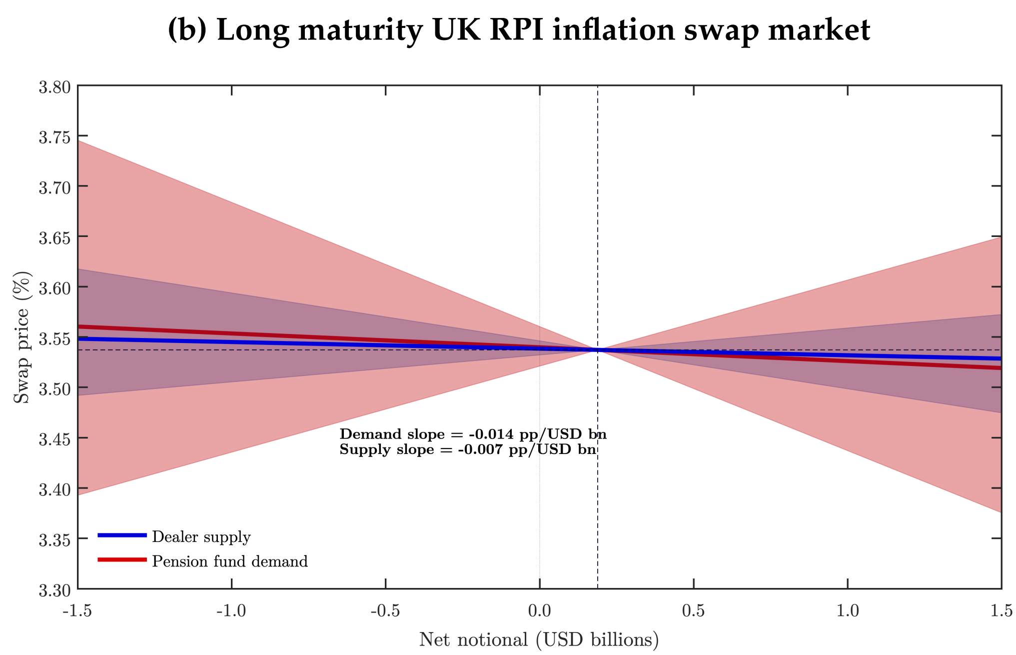
UK inflation swap prices experienced a period of heightened volatility during Autumn 2022. This was partly due to the knock-on effects of Russia's invasion of Ukraine, but this episode also came with a major crisis in the UK pension fund sector. Here we provide a narrative of events and describe how our model interprets the evolution of prices and quantities in the inflation swap market during this episode.

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 policy that would substantially reduce the effective prices that consumers would be paying for their household energy bills. This policy would have had 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 maturity bonds by liability-driven investment funds ("LDI" funds) 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 A9 shows long maturity inflation swap prices over the episode alongside the decomposition into frictions and expectations. Prices were unusually volatile during this period. The initial energy price guarantee brought an immediate decline in long dated inflation swaps, this is assigned to a change in expectations in our model. The fiscal expansion caused a sharp rise in inflation swaps initially, the announcement was deemed inflationary for fundamental reasons. 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 reversed, expected inflation fell by approximately

Figure A7 ESTIMATED SLOPES OF MARKET DEMAND AND SUPPLY FUNCTIONS USING GIV STRATEGY



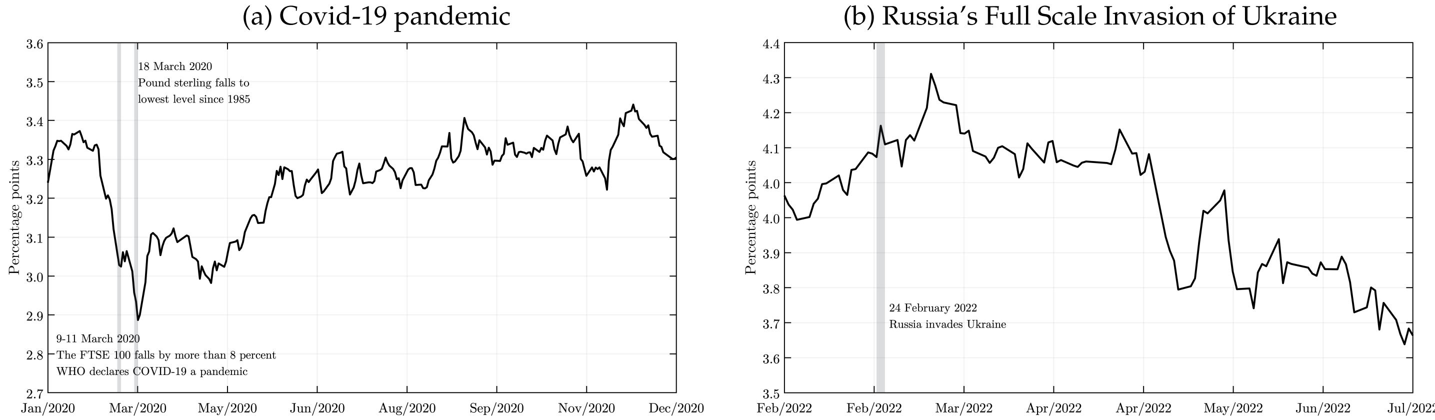


NOTE: Panel (a) estimates shows estimates of supply and demand curve for the short market using the instrumental variable strategy. Slope of the supply curve (blue) is defined as $\frac{dP/d\varepsilon_b}{dQ/d\varepsilon_b}$ using the impulse response functions. The slope of the demand curve (red) is defined as $\frac{dP/d\varepsilon_h}{dQ/d\varepsilon_h}$. The curves have intercepts such that P and Q are at the sample averages. The median estimate of the slopes from the sampler alongside 68% credible intervals are presented. Panel (b) repeats the exercise for the long market (i.e. p and q).

20bps and stayed persistently lower. After the dust settled, this combination of shocks lowered long maturity expected inflation.

Frictional shocks pushed swap prices higher by more than 10bps over the course of the crisis, obscuring the fall in expected inflation. This rise in liquidity premia happened mainly as the crisis was in its second week. This rise in liquidity premia largely reflects a contraction in supply by dealer banks and could reflect

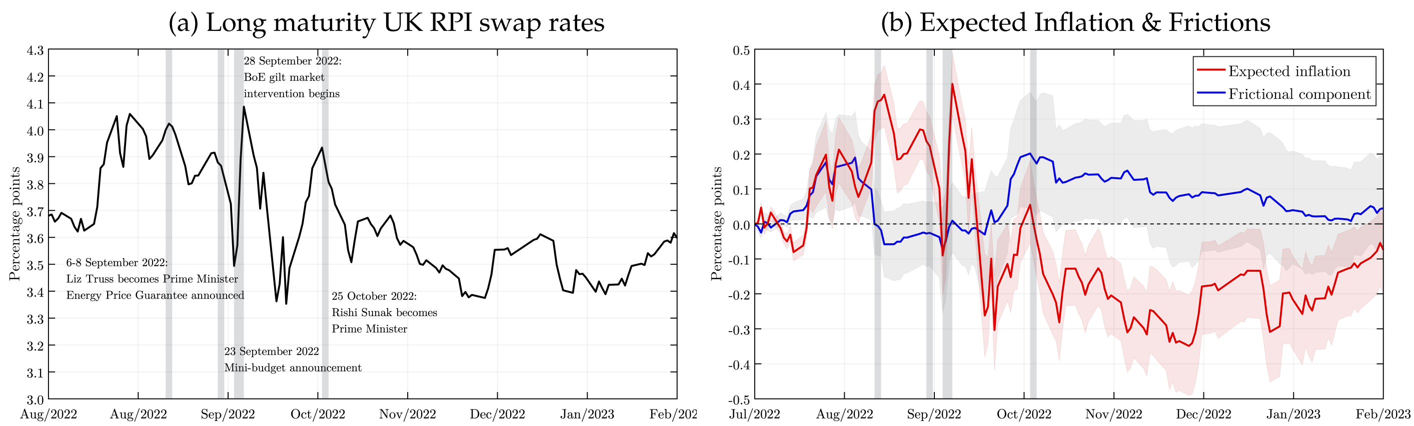
Figure A8 Two episodes: actual prices



Note: Weighted-average long maturity inflation swap breakeven rates, p, over the sample period for the Covid-19 episode (early 2020) and the initial phase of Russia's full-scale invasion of Ukraine (early February 2022).

concerns about the health of their counterparties within the pension fund sector that were unveiled as the crisis progressed. Earlier, in September, the dominant liquidity shock came from the demand side, and pushed prices down. According to our estimates, pension funds 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.

Figure A9 Autumn 2022: swap rates, liquidity, and expected inflation

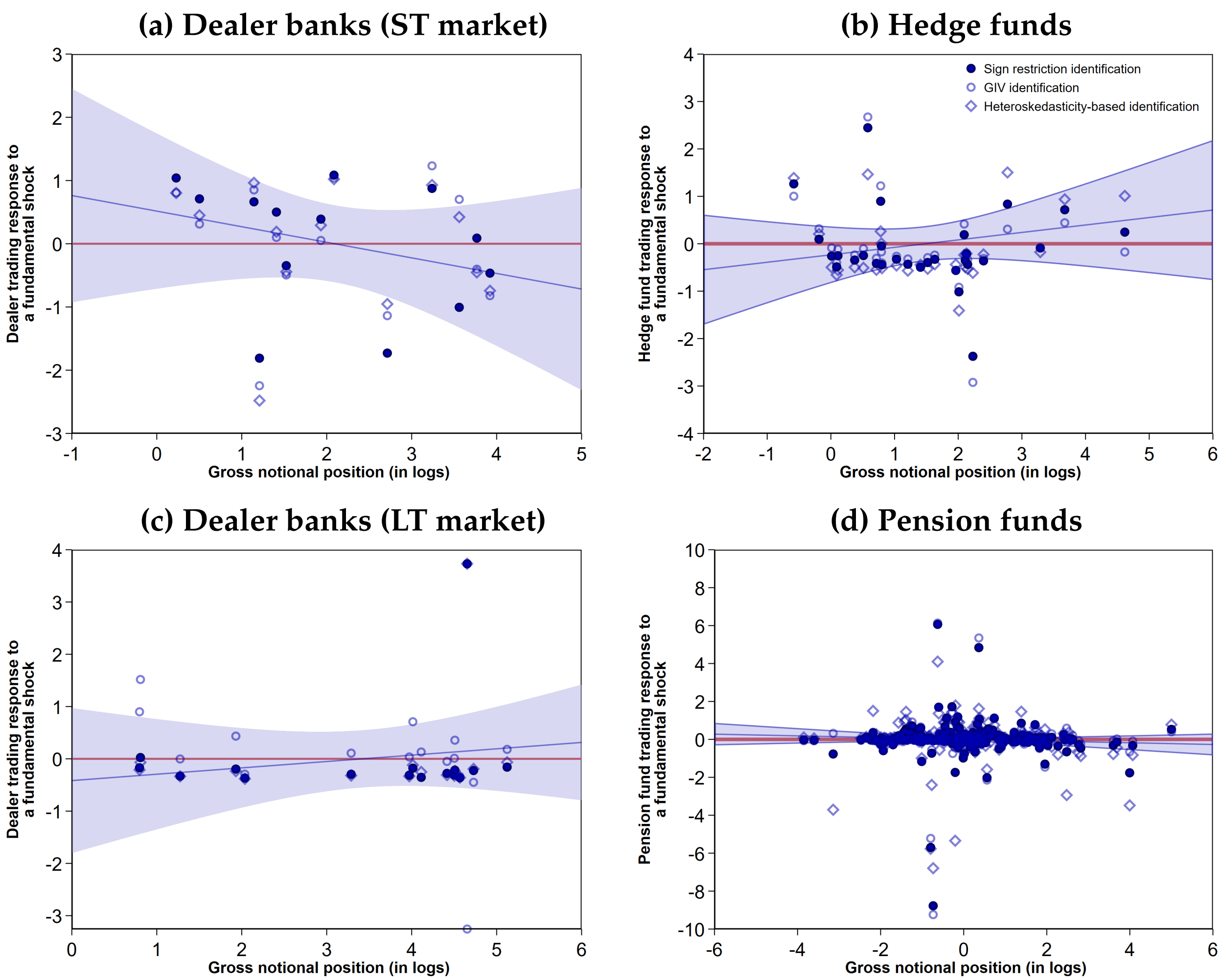


Note: Panel (a) shows the actual inflation breakeven rate, p, in the long maturity market that is a notional-weighted average of maturities longer than 10 years, over the period August 2022 until February 2023. Panel (b) shows the estimated cumulative contribution of inflation shocks, ε_π, and all remaining shocks to this price series. Estimates produced using the sign restrictions identification strategy. Solid lines are the median estimate from the sampler with shaded areas representing 68% credible intervals. The initial condition means the red and blue lines in panel (b) do not precisely sum to the cumulative change of the series plotted in panel (a).

D.4 Trade sensitivity and size of an institution

Figure A10 shows the scatter plots of the estimates β_i against the size of the institution in both the short maturity and long maturity market segments. There is little relation between the two.

Figure A10 COEFFICIENTS SCATTERED AGAINST SIZE OF INSTITUTION



NOTE: The figures are scatterplots of estimates of $\beta_{q,i}$, for $q \in \{b, h, f\}$ (we estimate $\beta_{b,i}$ for dealer banks trading in both the short maturity and long maturity markets) against institutional size $a_{q,i}$. Institutional size is proxied using the (log of) gross notional positions traded by each institution i that remains in effect in the data sample. The estimates of $\beta_{q,i}$ are obtained using the identified inflation shock from each of the three identification strategies as an instrument for the change in the swap breakeven rates in the respective market. The fitted line refers to estimates using sign restrictions, with 90% bootstrapped confidence intervals from 10,000 pseudo-samples drawing the relevant institutions in each sector with replacement to re-estimate the two data series. There are 30 hedge funds and 13 dealer banks n the short maturity market, and 210 pension funds and 16 dealer banks in the long maturity market. SOURCE: DTCC Trade Repository and authors' calculations.