CONVENIENCE YIELDS IN THE TIPS-TREASURY BOND MISPRICING

ELVIS HÖJMAN

TURE LJUNGBERG

Bachelor Thesis

Stockholm School of Economics

2026



Convenience Yields in the TIPS-Treasury Bond Mispricing

Abstract:

The TIPS-Treasury bond puzzle, documented by Fleckenstein, Longstaff, and Lustig (2014), describes a persistent and economically significant violation of the no-arbitrage relationship between U.S. nominal Treasury securities, Treasury Inflation-Protected Securities (TIPS), and inflation swaps. We empirically test an explanation that has been suggested but not yet studied directly: that the convenience yield of nominal Treasury securities accounts for the mispricing. Following van Binsbergen, Diamond, and Grotteria (2022), we measure the convenience yield as the spread between box rates implied by S&P 500 options and nominal Treasury yields, and we incorporate it into the no-arbitrage relationship. Using daily data from June 2007 to August 2025 at 1-, 2-, and 3-year tenors, we find that adjusting for the convenience yield reduces the average absolute deviation from no-arbitrage conditions by 5 to 10 basis points per annum and eliminates the directional bias previously documented in the literature. The convenience yield does not, however, explain time-series variation in the mispricing outside of the Great Financial Crisis. We interpret these findings as evidence that the convenience yield resolves the puzzle's most distinctive feature – its consistent direction – but acts primarily as a level shifter rather than a driver of high-frequency variation. Our results support including convenience yields alongside TIPS liquidity premia, Treasury credit risk, and intermediation frictions in any complete account of the puzzle.

Keywords:

TIPS-Treasury bond puzzle; convenience yield; box rates; inflation swaps; risk-free rate

Authors:

Elvis Höjman (25916)
Ture Ljungberg (25915)

Tutor:

Mehran Ebrahimian, Assistant Professor, Department of Finance

Examiner:

Ramin Baghai, Professor, Department of Finance

Bachelor Thesis
Bachelor Program in Business and Economics
Stockholm School of Economics
© Elvis Höjman and Ture Ljungberg, 2026

I. Introduction

In finance, few principles are held in as high regard as no-arbitrage. While researchers tend to accept small deviations from this principle, we sometimes uncover deviations so large that we must raise questions. One such case is what Fleckenstein et al. (2014) have termed "the TIPS-Treasury bond puzzle". They find consistent violations of a no-arbitrage relationship between nominal U.S. Treasury securities, U.S. inflation swaps – securities that give investors an inflation-indexed interest rate in exchange for a pre-determined, fixed interest rate –, and U.S. Treasury Inflation-Protected Securities (TIPS) – bonds that protect investors from inflation through having their principal indexed to the level of the Consumer Price Index (CPI). Haubrich, Pennacchi, and Ritchken (2012) express the no-arbitrage relationship in terms of yields[1]:

$$r_{t,n}^{inf} = r_{t,n}^{TSY} - r_{t,n}^{TIPS} \tag{1}$$

Where $r_{t,n}^{inf}$ is the fixed interest rate of inflation swaps, $r_{t,n}^{TSY}$ is the yield to maturity of nominal Treasury securities, and $r_{t,n}^{TIPS}$ is the real yield of TIPS, for continuously compounded zero-coupon interest rates of tenor n observed at time t.

Violations of this relationship are economically significant – averaging circa 50 basis points per annum – and consistent in direction, such that the spread between nominal Treasury yields and TIPS yields is consistently too small relative to inflation swap rates (Fleckenstein et al., 2014). With evidence that it is not mispricing of inflation swaps that drives the arbitrage (Fleckenstein et al., 2014; Ashton, 2008), one of two conditions must be true: TIPS are severely underpriced or nominal U.S. Treasuries are severely overpriced, to the tune of billions of dollars of total mispricing (Fleckenstein et al., 2014). It also means that there must exist a bias in at least one of the two most frequently used measures of U.S. risk-free rates – either the TIPS yield is a biased measure of the real risk-free rate, or the nominal Treasury yield is a biased measure of the nominal risk-free rate. A resolution to the TIPS-Treasury bond puzzle can thus enhance our understanding of how safe assets are priced and how we should measure risk-free rates in the American economy.

Researchers have searched for market frictions that can explain the TIPS-Treasury bond puzzle, such as liquidity premia in TIPS (D'Amico, Kim, and Wei, 2018; Andreasen, Christensen, and Riddell, 2021; Driessen, Nijman, and Simon, 2017; Pflueger and Viceira, 2016), credit risk in the U.S. Treasury (Dittmar, Hsu, Roussellet, and Simasek, 2026), intermediary frictions, and marking-to-market concerns (Ahn and Ahn, 2023). This paper investigates an explanation that existing literature has suggested, but that has not yet been rigorously empirically studied, namely that the convenience yield of nominal Treasury securities can explain the mispricing (Fleckenstein et al., 2014; Pflueger and Viceira, 2016). We investigate the following research questions:

1. Does the level of arbitrage violations between nominal Treasury securities, TIPS, and inflation swaps decrease when accounting for the convenience yield of nominal Treasury securities?

[1]The yield-relation specified by Haubrich et al. (2012) holds under the condition of frictionless markets.

2. Can variation in the arbitrage between nominal Treasury securities, TIPS, and inflation swaps be explained by variation in the convenience yield of nominal Treasury securities?

The concept of convenience yields, brought to prominence by Krishnamurthy and Vissing-Jorgensen (2012), is the finding that highly liquid safe assets like nominal U.S. Treasury securities have non-pecuniary benefits beyond the value of their cash flows – such as being widely accepted collateral and allowing financial institutions to satisfy capital requirements. Investors are willing to pay an additional price for these special properties, which lowers their yield below the true nominal risk-free rate. The purpose of this study is to test whether accounting for the price of these non-pecuniary benefits resolves the frequently documented TIPS-Treasury bond puzzle.

We draw on a different body of literature that measures the convenience yield in nominal Treasury securities by comparing their yields to the interest rate that options markets use to price options contracts, called the box rate. Following van Binsbergen et al. (2022) and Diamond and van Tassel (2022), we define the convenience yield as the spread between box rates and nominal Treasury yields. With this, we define two measures of the mispricing: First, the non-adjusted mispricing, which is the inflation swap rate less the spread between the nominal Treasury yield and the TIPS yield. Second, the mispricing adjusted for convenience yields, which is simply the non-adjusted mispricing less the convenience yield.[2] To investigate whether the level of arbitrage violations can be reduced by accounting for the convenience yield, we test whether the absolute value of the mispricing is smaller after adjusting for the convenience yield on average. Since the mispricing should be zero in a no-arbitrage market, any deviation from zero constitutes a mispricing. Further, we investigate whether variation in the convenience yield can explain variation in the mispricing, we estimate an ordinary least squares (OLS) regression with the non-adjusted mispricing as the dependent variable and the convenience yield as the independent variable. We conduct these tests both over our entire sample period spanning from 2007 to 2025, and over a subsample excluding the Great Financial Crisis (GFC) to see if our results are persistent over time or simply driven by unique dynamics in crisis periods.

We find that adjusting for the convenience yield statistically significantly decreases the average absolute deviation from no-arbitrage conditions by between 5 and 10 basis points per annum in the full sample and between 4 and 7 basis points in the subsample excluding the GFC, bringing markets significantly closer to satisfying no-arbitrage conditions. Further, the previously documented consistent overpricing of Treasury securities relative to TIPS no longer exists after adjusting for convenience yields. Further, while we find that the convenience yield explains between 18 and 35 percent of the variation in the mispricing over our entire study period, we note that this explanatory power is practically entirely driven by the GFC and that the explanatory power does not remain after excluding it.

Our results consistently show that the convenience yield is of the right size and direction to resolve much of the observed mispricing in the TIPS-Treasury bond puzzle, but that it is not a meaningful predictor of the mispricing outside of crisis episodes. Our interpretation is that when the convenience yield varies strongly, such as during the GFC, its effect on the TIPS-Treasury bond puzzle can be detected. During normal times, the convenience yield is small in

[2] An explanation of how the convenience yield should enter the yield relationship is presented in Section III.

magnitude and its variation is dominated by noise in estimation, making it a poor statistical predictor in the observed mispricing, although a valuable mean-shifter in the observed arbitrage puzzle. A more skeptical interpretation would suggest that the literature is wrong in suggesting the convenience yield as a substantial factor in the TIPS-Treasury bond puzzle, and that the economic significance of the convenience yield in this context is negligible.

This paper proceeds as follows: In Section II, we shortly outline our contributions to the current literature. In Section III, we provide background on the convenience yield: What it is, how it should enter the studied pricing relationship, and why it can be measured as the spread between box rates and U.S. Treasury yields. In Section IV, our data is described. In Section V, we outline our methodology. Our results are presented in Section VI with robustness and diagnostic tests presented in Section VII. Finally, the results are discussed in Section VIII and concluded in Section IX.

II. Related Literature

This paper mainly contributes to two bodies of literature. We most directly contribute to the literature on the TIPS-Treasury bond puzzle (Fleckenstein et al., 2014; Andreasen et al., 2021; Driessen et al., 2017; Pflueger and Viceira, 2016; Dittmar et al., 2026; Ahn and Ahn, 2023). Most existing empirical analyses have argued that the explanation for the arbitrage between nominal Treasury securities, TIPS, and inflation swaps is the relative illiquidity of TIPS compared to nominal Treasury securities (D'Amico et al., 2018; Andreasen et al., 2021; Driessen et al., 2017; Pflueger and Viceira, 2016). This research is generally in agreement that TIPS liquidity explains a meaningful part, but not all of the mispricing, with Driessen et al. (2017) providing a reasonable range of between 29 percent and 63 percent being explained by TIPS illiquidity. Newer research finds different drivers of the mispricing. Dittmar et al. (2026) find that U.S. Treasury credit risk explains between 46 percent and 78 percent of the mispricing depending on tenor. Ahn and Ahn (2023) find that marking-to-market concerns explain around 53 percent of the mispricing and that marking-to-market concerns and intermediation frictions combined explain around 70 percent. Despite these advances, no single channel has emerged as a definitive explanation. We contribute to this literature by providing the first direct empirical test of the conjecture, originally proposed by Fleckenstein et al. (2014) and tentatively echoed by Pflueger and Viceira (2016), that the convenience yield in nominal Treasury securities is responsible for the arbitrage. We find that accounting for convenience yields consistently brings the mispricing closer to zero, consistent with their hypothesis, but that this effect operates primarily as a level adjustment: The average convenience yield is of approximately the right size to neutralize the documented directional bias, while variation in the convenience yield is only informative about variation in the mispricing during crisis episodes.

Further, we contribute to the body of literature that tries to identify the optimal reference rate to measure convenience yield with. A large body of literature uses a variety of ways to measure the convenience yield, including corporate bond spreads (Krishnamurthy and Vissing-Jorgensen, 2012), near money assets (Nagel, 2016), Refcorp bonds (Longstaff, 2004), CIP-implied foreign benchmarks (Du, Im, and Schreger, 2018), and box rates (van Binsbergen et

al., 2022; Diamond and van Tassel, 2022). Although the literature agrees that an economically significant convenience yield in nominal Treasury securities exists, there is a lack of consensus around the best way to measure this convenience yield with. van Binsbergen et al. (2022) and Diamond and van Tassel (2022) provide a compelling theoretical argument that the box rate is the best reference rate since it is the only rate that is both risk-free and carries no convenience yield. A good way to test if the box rate is a valuable reference rate is to test if using it can explain known anomalies in finance like covered interest parity (CIP) deviations (van Binsbergen et al., 2022; Diamond and van Tassel, 2022). This paper contributes to the field by finding evidence that box rates can indeed be used to resolve some part of the TIPS-Treasury bond puzzle, suggesting their economic validity as a reference point with which to measure the convenience yield. This being said we also clearly display the main limitation of using box rates, namely substantial noise and volatility in the estimated interest rates.

III. Convenience Yields

A. What is the Convenience Yield?

The convenience yield is the non-pecuniary value that investors derive from holding certain safe assets, such as the safety, liquidity, and money-like services the safe asset provides. It manifests as a yield discount: investors accept a lower financial return on convenient assets because they value the implicit service flow. This convenience yield is especially pronounced in nominal Treasury securities – the nominal Treasury rate is consistently below benchmarks of other 'pure' risk-free rates (Krishnamurthy and Vissing-Jorgensen, 2012; van Binsbergen et al., 2022). Further, the convenience yield is persistent and cannot be arbitraged away by shorting the nominal Treasury security and going long the 'pure' risk-free asset since the non-pecuniary benefits of the Treasury security cannot be replicated by the synthetic long leg. The lender of the Treasury temporarily loses the asset's convenience and subsequently demands compensation for the loss of non-pecuniary value, thus eliminating the arbitrage opportunity in equilibrium (Krishnamurthy, 2002).

B. How Should the Convenience Yield Enter the TIPS-Treasury Bond Puzzle?

The convenience yield becomes directly relevant for Haubrich et al.'s (2012) no-arbitrage relationship between the yields of nominal Treasury securities, TIPS, and inflation swaps. This relationship does not account for the fact that investors are willing to pay for non-pecuniary benefits. While TIPS and inflation swaps can be used to replicate the cash flows of a nominal Treasury security, they cannot replicate the convenience of said Treasury security. Equivalently expressed, TIPS and inflation swaps can only be used to replicate the cash flows of a risk-free asset with no convenience. Haubrich et al.'s (2012) yield-based relationship can thus be respecified as:

$$r_{t,n}^{inf} = r_{t,n}^{risk\ free} - r_{t,n}^{TIPS} \tag{2}$$

Where $r_{t,n}^{risk\ free}$ is the continuously compounded risk-free rate of tenor n observed at time t, free of any convenience yield.

Assuming that the wedge between the 'pure' nominal risk-free rate and the nominal Treasury yield is composed of only the convenience yield, equation (2) can be re-expressed as our preferred specification of the no-arbitrage relationship:

$$r_{t,n}^{inf} = (r_{t,n}^{TSY} + r_{t,n}^{conv}) - r_{t,n}^{TIPS} \tag{3}$$

Where $r_{t,n}^{conv}$ is the continuously compounded convenience yield of zero-coupon nominal Treasury securities of tenor n observed at time t.

Although equation (2) and (3) are equivalent under our assumptions, we prefer the specification which makes explicit the convenience yield instead of using equation (2) directly, for two reasons: (i) to highlight the role of extra-monetary factors in resolving the TIPS-Treasury bond puzzle, and (ii) to make the time-varying structure of the convenience yield and its effect on the TIPS-Treasury bond puzzle directly observable.

C. Why the Spread Between the Box Rate and the Nominal Treasury Yield is an Adequate Measure of the Convenience Yield

A box trade on European options is risk-free, providing a guaranteed future payoff in exchange for a fixed price today. At the same time, options contracts are risky assets, meaning they do not carry a convenience yield. Thus, the interest rate one can gain through a box trade – the so-called box rate – is a risk-free interest rate that is free of any convenience yield. This can be exploited to measure the convenience yield of nominal Treasury securities as the spread between the box rate and the yield of nominal Treasury securities. Such a measure is superior to other often-used measures of the convenience yield in two ways: (i) it avoids measuring spreads between two safe assets, in which case we would only get a measure of the difference in convenience yields, and (ii) it avoids measuring spreads using assets with credit risk, in which case the convenience yield measure is confounded by a credit-risk premium (van Binsbergen et al., 2022; Diamond and van Tassel, 2022).

There are three main concerns that could render this measurement of the convenience yield invalid: A spread in credit risk between the issuer of European SPX options and the U.S. Treasury, a spread in liquidity premia between box trades and U.S. Treasury securities, and additional cost of capital in the box rate imposed by margin requirements for the box trade.

Credit risk in European SPX options is considered negligible by the precedent literature, due to the large institutional protections that exist on options markets for the following reasons (van Binsbergen et al., 2022): Any investor in SPX options must post daily variation margin to the Chicago Board Options Exchange (CBOE). Next, CBOE posts additional margin to the Options Clearing Corporation (OCC), which is designated a "Systematically Important Financial Market Utility" and subsequently has access to emergency liquidity support from the United States Federal Reserve. Thus, a default of an options contract would require (i) a default of the individual trader counterparty, (ii) a default of CBOE, and at the same time (iii) a default of OCC, with the United States Federal Reserve liquidity support being insufficient to avoid default. In any default event, options contracts are considered "super-senior", meaning traders have an immediate claim to the value of their contract before any debtor is paid.

Regarding liquidity premia in SPX options, research is very limited. However, research has found that there is a liquidity premium in exchange-traded equity options (Choy and Wei, 2020; Christoffersen, Goyenko, Jacobs, and Karoui 2018). In light of the fact that SPX options are substantially more liquid than equity options (Schwartz, 2026), the liquidity premium in SPX options should be a substantially smaller concern; however, we can not state with confidence that the effects of this are negligible. In Section VII.A we find that our box rates are meaningfully different to box rates constructed from options on a different index where contracts are less liquid. A more thorough investigation of this and its effects on estimated box rates is left for future research.

Lastly, it is worth noting that CBOE acknowledges the box trade as a riskless trade, eliminating margin constraints that otherwise could increase the cost of capital of executing a box trade to harvest the box rate (van Binsbergen et al., 2022).

For the Treasury security side: Credit risk in the U.S. Treasury is remote but non-zero, with a notable technical default event in 1979. Augustin, Chernov, Schmid, and Song (2021) argue that the credit-risk component of U.S. nominal Treasury yields is non-negligible in affecting benchmark-rate interpretations. Liquidity in the U.S. nominal Treasury security market is, in normal times, among the highest of any bond market globally, particularly for on-the-run securities (Fleming, 2003; Fleming, Mizrach, and Nguyen, 2018). Nonetheless, Treasury yields embed a measurable liquidity premium: Longstaff (2004) compares Treasuries to credit-equivalent Refcorp bonds and finds an economically significant flight-to-liquidity premium in stress periods, while Hu, Pan, and Wang (2013) document persistent yield-curve fitting noise that proxies for market-wide illiquidity. Thus, there may be factors beyond convenience that drive the yield of nominal Treasury securities to differ from the true risk-free rate.

However, deviations on the side of nominal Treasury securities should not devalue the model we suggest: If box rates are accepted as the true risk-free rate, then any wedge to U.S. Treasury bond rates – which we denote "the convenience yield" but which could also be composed of other factors – must be adjusted for. One can imagine that the spread between box rates and U.S. Treasury bond yields is composed of (i) the convenience yield, (ii) a liquidity premium, and (iii) a credit risk premium, but in such a case, the no-arbitrage relationship we study should also be adjusted for both (i), (ii) and (iii). Thus, the concern of unobservable factors biasing the nominal Treasury yield should only weaken the interpretation of our results, not the validity of our model. Our model can thus only be invalidated by discrepancies between box rates and the true risk-free rate.

IV. Data

The data used for this study were gathered from three sources: London Stock Exchange Group (LSEG) Workspace, Wharton Research Data Services (WRDS), and publicly available information from the Board of Governors of the Federal Reserve System (commonly known as the Federal Reserve Board). All data in our sample for our main analysis run from the first date our inflation swap data are available, June 18, 2007, to the last date that our options data are available, August 29, 2025.

Our main analyses are limited to 1-, 2-, and 3-year tenors due to the maturities at which SPX options are issued. The SPX options in our dataset have maturities up to 3 years, but not further. This means that Nelson-Siegel-Svensson (NSS)-estimation[3] to tenors longer than 3 years has no actual observations to anchor to, which, as argued by Gürkaynak, Sack, and Wright (2007), makes the estimation and the results thereof unstable.

In total, our sample with observations from each trading day from June 18, 2007 to August 29, 2025, after removing days with missing observations in any series, results in a total sample size of 4547 observations.

A. Nominal Treasury, TIPS, and Inflation Swap Data

For the construction of the basic mispricing time series, we first collect a daily multivariate time series of zero-coupon nominal U.S. Treasury yields from the Gürkaynak et al. (2007) dataset, maintained by the Federal Reserve Board. This dataset contains continuously compounded annual yields at 1-, 2-, and 3-year tenors. The yields are derived from an NSS curve fitted to off-the-run Treasury securities. Second, we retrieve a daily multivariate time series of zero-coupon TIPS yields from the Gürkaynak, Sack, and Wright (2010) dataset, maintained by the Federal Reserve Board. This dataset contains continuously compounded annual yields at 2- and 3-year tenors. Similar to the nominal yield data, the yields are derived from an NSS-estimation. The dataset also contains the estimated NSS-parameters which we use to extrapolate the 1-year tenor yield. It is, however, important to note that Gürkaynak et al. (2010) actively chose to not extrapolate that far since their initial dataset of TIPS prices only included tenors above 18 months. We thus note that interpretations at 1-year maturities must be made with caution. Lastly, we gathered zero-coupon U.S. inflation swap data from LSEG at 1-, 2-, and 3-year tenors. They provide aggregated daily time series of annually compounded annual swap rates based on data from a multitude of contributors, which we convert to a continuously compounded basis. The inflation swaps use CPI-U as the underlying series to match the measure used for indexation of TIPS.

Figures 1-3 present the continuously compounded annual yields for TIPS, nominal Treasury securities, and inflation swaps at 1-, 2-, and 3-year tenors respectively.

[3]For details on NSS-estimation, see Appendix A

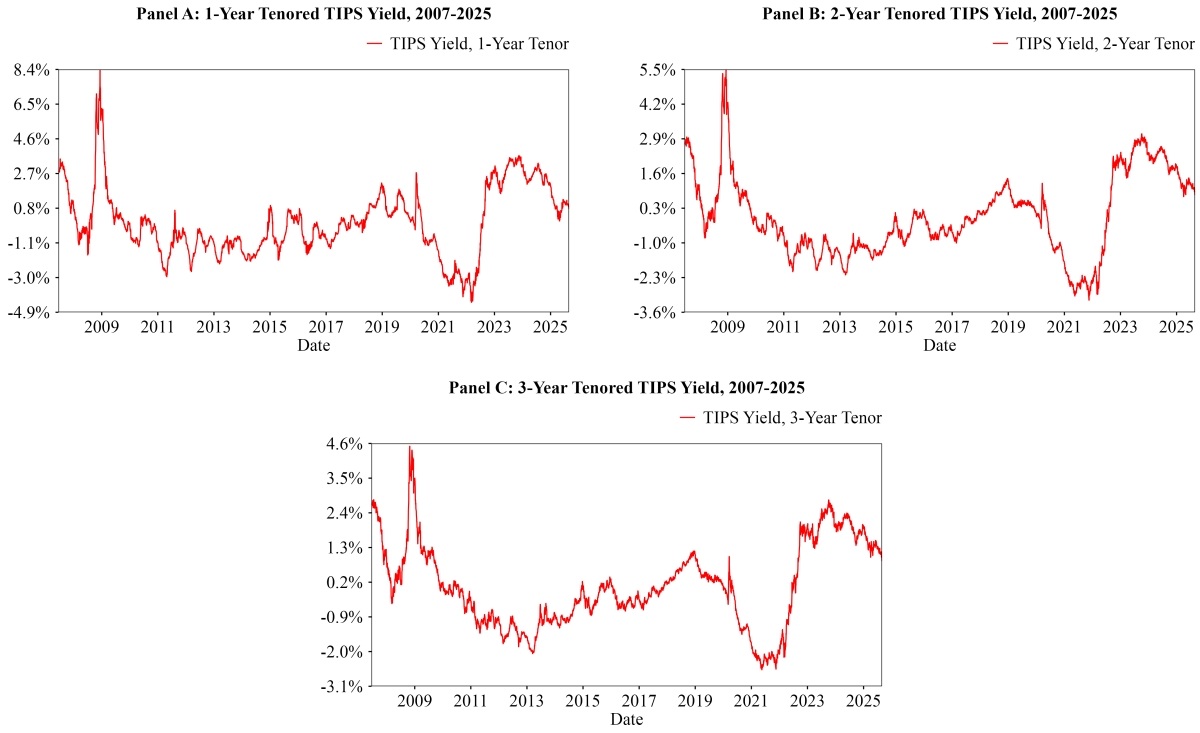


Figure 1. *Continuously Compounded Annual TIPS Yields, at 1-Year, 2-Year, and 3-Year Tenors*

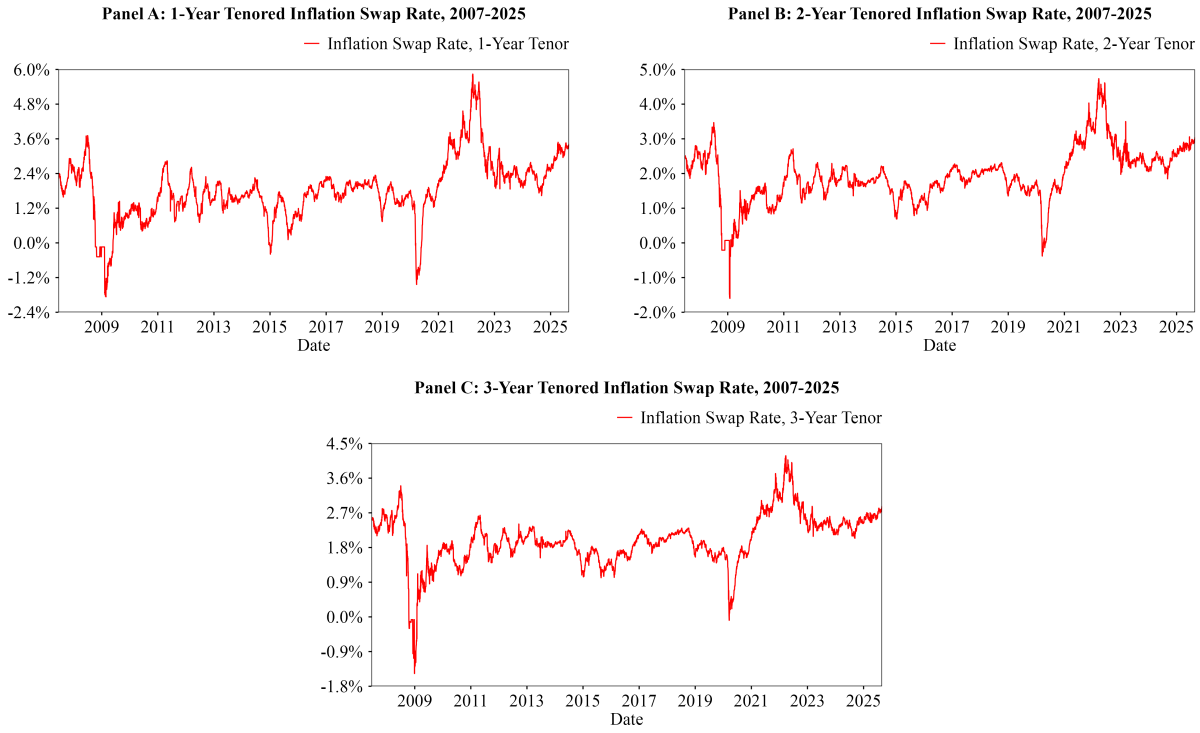


Figure 2. *Continuously Compounded Annual Interest Rates on U.S. CPI-U Inflation Swaps, at 1-Year, 2-Year, and 3-Year Tenors*

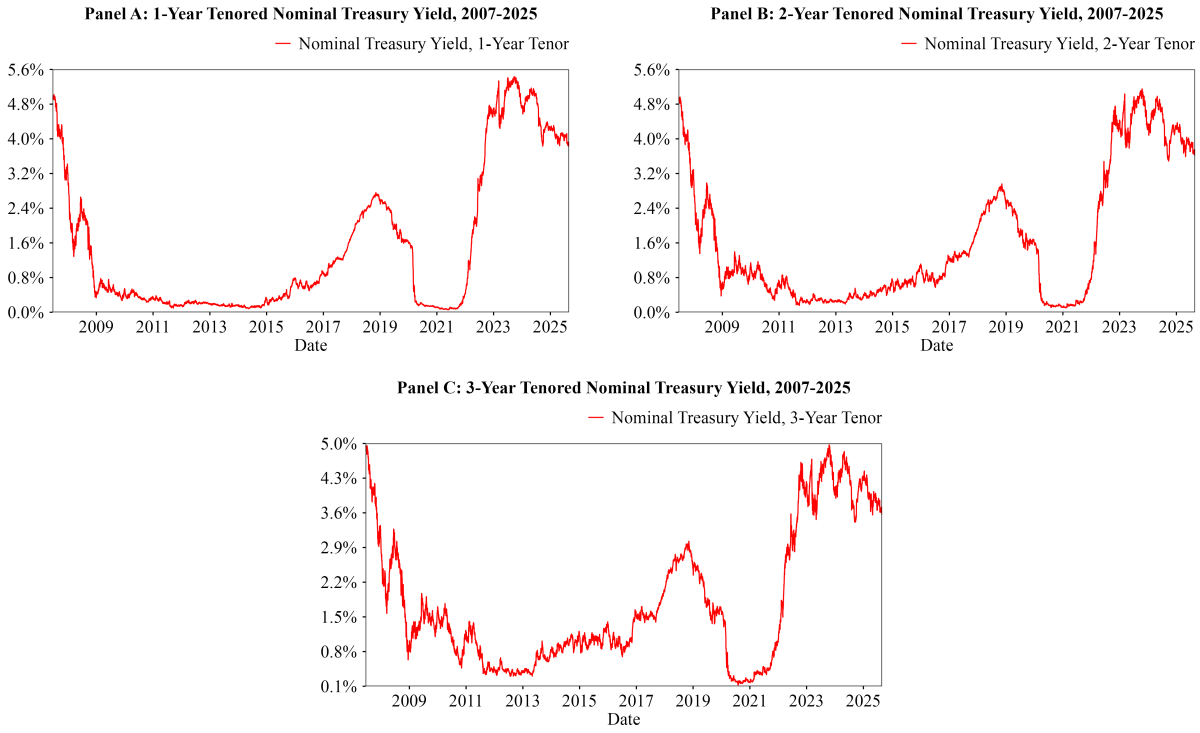


Figure 3. *Continuously Compounded Annual Nominal Treasury Yields, at 1-Year, 2-Year, and 3-Year Tenors*

In the TIPS yields, we see very similar general trends across maturities. We also note that the broad trend at the 1-year tenor that we extrapolated from the NSS-parameters in the Gürkaynak et al. (2010) dataset is consistent with the longer maturities, supporting the validity of our estimation. However, it is worth noting that the 1-year tenor exhibits greater volatility, as exemplified during the GFC when the 1-year tenor yield reached above 8 percent while the 2-year tenor yield only reached slightly above 5 percent. This is likely because relatively small variations in NSS parameters translate more directly to changes in the yield curve at shorter tenors, especially when there are no observed bonds at that tenor.

Comparing the different yields, we can see that TIPS yields and nominal Treasury yields follow similar broad trends, which is to be expected given that nominal yields consist partly of real yields. The most noticeable difference of trends between these two are around the GFC, where TIPS yields show a much larger spike than nominal yields. This is partly because of genuine disinflation expectations raising TIPS yields, but has also been explained by liquidity dislocation where investors fled away from less liquid securities like TIPS to more liquid ones like nominal Treasuries (Campbell, Shiller, and Viceira, 2009; D'Amico et al., 2018).

B. SPX Options Data

For the construction of the box rates, we follow van Binsbergen et al. (2022) and collect options data on the S&P 500 index (SPX), as these are the most liquid European options on the U.S. market. For robustness, we also collect options data on the Russell 2000 index (RUT) to test if estimated box rates materially differ depending on the choice of underlying index (our RUT data starts 2011-03-04). The options data is a daily panel dataset of European options on

the indices, observed at the end of each trading day. The data is a collection of all options at all traded expiration dates from the Chicago Board Options Exchange (CBOE) – who are the exclusive issuers of SPX and RUT options – collected by the Center for Research in Security Prices (CRSP). We access the dataset through WRDS. For the end of each trading day, the dataset contains the observation date, the expiration date, the strike price, the best ask price, and the best offer price for all puts and calls. Following van Binsbergen et al. (2022), we define the price of an options contract as the "mid price" – the mean of the best ask price and the best offer price. Days until expiry of an option is defined as the expiry date minus the observation date.

Tables 1-3 present the mean strike price, the strike price range, the mean option price, and the mean bid-ask spread as a percentage of the mean price, for puts and calls separately over 7-, 30-, and 365-day expiries.

Panel A: Calls

Year	Mean Strike	Strike Range	Mean Price	Mean Spread (%)
2007	1,452	[1,100; 1,600]	57.9	5.0%
2008	1,149	[450; 1,650]	61.0	5.0%
2009	925	[400; 1,650]	85.0	2.7%
2010	1,082	[400; 1,550]	85.3	2.5%
2011	1,211	[100; 2,000]	75.2	2.3%
2012	1,320	[400; 2,000]	96.0	2.0%
2013	1,548	[400; 2,075]	140.1	1.6%
2014	1,729	[400; 2,500]	241.7	2.3%
2015	1,781	[100; 3,500]	315.2	2.1%
2016	1,894	[100; 3,500]	251.3	2.6%
2017	2,231	[100; 3,500]	254.9	2.4%
2018	2,531	[100; 3,800]	279.7	1.3%
2019	2,650	[100; 4,800]	319.3	1.1%
2020	2,955	[100; 5,600]	383.9	0.8%
2021	3,900	[100; 7,000]	512.9	0.6%
2022	3,934	[100; 7,600]	326.8	1.0%
2023	4,008	[100; 8,300]	386.6	0.7%
2024	5,037	[200; 10,000]	511.3	1.1%
2025	5,605	[200; 12,000]	484.0	1.4%

Panel B: Puts

Year	Mean Strike	Strike Range	Mean Price	Mean Spread (%)
2007	1,452	[1,100; 1,600]	15.4	11.9%
2008	1,149	[450; 1,650]	57.2	4.7%
2009	924	[400; 1,650]	54.2	3.2%
2010	1,082	[400; 1,550]	30.2	4.8%
2011	1,211	[100; 2,000]	30.5	4.2%
2012	1,320	[400; 2,000]	27.9	3.7%
2013	1,548	[400; 2,075]	21.6	3.7%
2014	1,729	[400; 2,500]	21.0	5.5%
2015	1,781	[100; 3,500]	38.1	5.3%
2016	1,894	[100; 3,500]	35.8	5.6%
2017	2,231	[100; 3,500]	30.4	6.5%
2018	2,531	[100; 3,800]	68.7	3.2%
2019	2,650	[100; 4,800]	57.6	2.5%
2020	2,955	[100; 5,600]	140.6	1.6%
2021	3,900	[100; 7,000]	117.2	1.2%
2022	3,934	[100; 7,600]	179.2	1.5%
2023	4,008	[100; 8,300]	100.2	1.3%
2024	5,037	[200; 10,000]	99.1	2.3%
2025	5,605	[200; 12,000]	161.5	2.1%

Table 1. *Summary Statistics of 7-Day European SPX Options*

Note: All values in USD. Price is the mid price of the options. Spread is the bid-ask spread divided by the mid price. 2007 values are calculated starting 2007-06-18. 2025 values are calculated ending 2025-08-29. Means are not volume-weighted.

Panel A: Calls

Year	Mean Strike	Strike Range	Mean Price	Mean Spread (%)
2007	1,407	[600; 1,800]	139.2	1.3%
2008	1,193	[200; 2,000]	90.6	2.5%
2009	894	[200; 2,500]	152.1	1.3%
2010	1,003	[100; 2,500]	193.4	1.4%
2011	1,106	[50; 3,000]	213.2	1.1%
2012	1,218	[100; 3,000]	204.9	1.2%
2013	1,480	[100; 3,000]	201.5	1.0%
2014	1,702	[100; 3,000]	268.3	1.5%
2015	1,780	[100; 3,500]	319.3	1.8%
2016	1,821	[100; 3,500]	313.3	1.8%
2017	2,209	[100; 3,500]	295.7	1.6%
2018	2,494	[100; 4,000]	322.0	1.2%
2019	2,612	[100; 4,800]	365.6	0.9%
2020	2,885	[100; 5,500]	469.6	0.9%
2021	3,783	[100; 7,200]	615.5	0.5%
2022	3,924	[100; 7,600]	407.7	1.0%
2023	3,949	[100; 8,300]	490.6	0.8%
2024	4,934	[200; 10,000]	638.4	1.0%
2025	5,539	[200; 12,000]	639.6	1.1%

Panel B: Puts

Year	Mean Strike	Strike Range	Mean Price	Mean Spread (%)
2007	1,407	[600; 1,800]	38.0	3.7%
2008	1,193	[200; 2,000]	136.5	2.3%
2009	894	[200; 2,500]	86.8	1.8%
2010	1,003	[100; 2,500]	56.0	2.8%
2011	1,106	[50; 3,000]	62.7	2.7%
2012	1,218	[100; 3,000]	41.2	3.3%
2013	1,480	[100; 3,000]	31.6	3.0%
2014	1,704	[100; 3,000]	27.3	4.4%
2015	1,781	[100; 3,500]	42.4	4.4%
2016	1,821	[100; 3,500]	44.8	3.9%
2017	2,209	[100; 3,500]	35.7	4.6%
2018	2,494	[100; 4,000]	72.9	3.0%
2019	2,612	[100; 4,800]	66.9	2.1%
2020	2,885	[100; 5,500]	148.8	2.1%
2021	3,783	[100; 7,200]	121.4	1.2%
2022	3,924	[100; 7,600]	242.5	1.4%
2023	3,949	[100; 8,300]	131.5	1.2%
2024	4,934	[200; 10,000]	125.0	1.9%
2025	5,539	[200; 12,000]	210.7	1.8%

Table 2. *Summary Statistics of 30-Day European SPX Options*

Note: All values in USD. Price is the mid price of the options. Spread is the bid-ask spread divided by the mid price. 2007 values are calculated starting 2007-06-18. 2025 values are calculated ending 2025-08-29. Means are not volume-weighted.

Panel A: Calls

Year	Mean Strike	Strike Range	Mean Price	Mean Spread (%)
2007	1,380	[600; 2,000]	247.6	1.2%
2008	1,231	[200; 2,500]	131.3	2.2%
2009	1,059	[100; 2,500]	202.9	1.5%
2010	983	[50; 2,500]	345.6	1.3%
2011	1,124	[100; 3,000]	280.0	1.8%
2012	1,183	[100; 3,000]	326.3	1.6%
2013	1,256	[100; 3,000]	455.6	0.7%
2014	1,478	[100; 3,000]	567.2	0.8%
2015	1,612	[100; 3,500]	526.8	1.3%
2016	1,713	[100; 3,500]	521.5	1.1%
2017	1,893	[100; 3,500]	683.4	1.0%
2018	2,284	[100; 4,200]	600.9	1.3%
2019	2,454	[100; 4,600]	669.6	0.9%
2020	2,818	[100; 6,500]	742.7	4.2%
2021	3,759	[100; 7,600]	981.0	1.5%
2022	3,748	[100; 8,200]	835.9	1.4%
2023	3,848	[200; 9,200]	918.4	0.9%
2024	4,836	[200; 12,000]	1,208.4	4.3%
2025	5,310	[200; 12,000]	1,259.9	1.0%

Panel B: Puts

Year	Mean Strike	Strike Range	Mean Price	Mean Spread (%)
2007	1,380	[600; 2,000]	92.8	3.0%
2008	1,231	[200; 2,500]	315.5	1.2%
2009	1,059	[100; 2,500]	240.8	1.2%
2010	983	[50; 2,500]	163.2	2.3%
2011	1,124	[100; 3,000]	197.9	2.3%
2012	1,183	[100; 3,000]	151.2	2.9%
2013	1,256	[100; 3,000]	112.0	2.1%
2014	1,482	[100; 3,000]	91.8	3.8%
2015	1,612	[100; 3,500]	131.5	3.2%
2016	1,713	[100; 3,500]	165.5	1.7%
2017	1,893	[100; 3,500]	98.3	2.3%
2018	2,284	[100; 4,200]	180.5	2.4%
2019	2,454	[100; 4,600]	169.6	1.4%
2020	2,818	[100; 6,500]	363.7	7.1%
2021	3,759	[100; 7,600]	443.1	2.9%
2022	3,748	[100; 8,200]	466.9	1.8%
2023	3,848	[200; 9,200]	333.5	1.1%
2024	4,836	[200; 12,000]	355.4	8.2%
2025	5,310	[200; 12,000]	452.6	1.3%

Table 3. *Summary Statistics of 365-Day European SPX Options*

Note: All values in USD. Price is the mid price of the options. Spread is the bid-ask spread divided by the mid price. 2007 values are calculated starting 2007-06-18. 2025 values are calculated ending 2025-08-29. Means are not volume-weighted.

The first thing to note is that the mean strike price and strike price range align for puts and calls across maturities, since CBOE issues puts and calls on the same strike price ladder. This is beneficial since our methodology requires the matching of puts and calls by strike price.

Since our model makes the simplifying assumption that the price at which options are traded is the mid price, the transaction prices that we use contain measurement errors that are bounded

by the bid-ask spread. Under the assumption that any difference between the mid price and the true transaction price is uncorrelated with the strike price of the option, this introduces random measurement error in the dependent variable of our box rate regression[4], which should not bias our estimates of the box rates but which may introduce noise. We do note that the bid-ask spread is generally higher in times of high market uncertainty.

V. Methodology

The no-arbitrage relationship between inflation swaps, nominal Treasury yields, and TIPS yields is defined as the inflation swap rate being equal to the spread between the nominal Treasury yield and the TIPS yield (see equation 1). Although no-arbitrage relations are commonly expressed in terms of cash flows, this yield-based relationship is an exact no-arbitrage relationship for continuously compounded rates of zero-coupon securities under the condition of frictionless markets (Haubrich et al., 2012). From this, we define the observed mispricing as:

$$A^1_{t,n} = r^{inf}_{t,n} - \left[r^{TSY}_{t,n} - r^{TIPS}_{t,n}\right] \tag{4}$$

The main contribution of this paper is to re-specify the no-arbitrage relationship between TIPS, nominal Treasury securities, and inflation swaps to include the convenience yield of nominal Treasuries as in equation (3). From this, we re-define the observed mispricing as:

$$A^2_{t,n} = r^{inf}_{t,n} - \left[(r^{TSY}_{t,n} + r^{conv}_{t,n}) - r^{TIPS}_{t,n}\right]$$

Or equivalently:

$$A^2_{t,n} = A^1_{t,n} - r^{conv}_{t,n}$$

We follow the methodology of van Binsbergen et al. (2022) to define the convenience yield of nominal Treasury securities as:

$$r^{conv}_{t,n} = r^{box}_{t,n} - r^{TSY}_{t,n}$$

Where r^{box} is the continuously compounded annual interest rate one can gain by executing a box trade in European options on the SPX.

To answer research question 1 – if accounting for the convenience yield reduces the mispricing – we compare A^1 to A^2 in absolute value over our test period by performing a one-sided Newey-West test of mean difference:

$$D_{t,n} = \left|A^1_{t,n}\right| - \left|A^2_{t,n}\right|$$

This tests whether the arbitrage violation is reduced on average when accounting for the convenience yield of nominal Treasuries. Since both a mispricing below zero and above zero constitutes a mispricing, we conduct this test on absolute values. If accounting for the convenience yield does reduce the mispricing, we would expect the average level of D to be statistically significantly greater than zero.

[4]The box-rate regression as specified in equation (5), introduced under Section V.A.

To answer research question 2 – if variation in the convenience yield can explain variation in the mispricing – we test whether the mispricing A^1 significantly co-moves with the convenience yield of nominal Treasury securities by defining the regression:

$$A^1_{t,n} = \alpha + \beta \times r^{conv}_{t,n} + e_{t,n}$$

If the convenience yield can explain variation in the mispricing, we would expect β to be statistically significantly different from zero.

Throughout this paper, we report p-values based on pre-whitened Newey-West standard errors with dynamic bandwidth selection, to account for variation in the length of autocorrelation throughout the sample period.

To conduct these tests, we need a multivariate time series of r^{inf}, r^{TSY}, r^{box}, and r^{TIPS} at all selected tenors. Time series of r^{inf} and r^{TSY} are directly imported from the data described under Section IV. The method for creating time series of r^{box} and subsequently r^{conv} is described under Section V.A. A seasonal adjustment is applied to A^1, to account for an artefact introduced by the NSS-estimation of TIPS yields. The motivation behind and implementation of the seasonal adjustment is described under Section V.B. All tests are executed over two distinct samples: (i) our entire sample period, and (ii) over a trimmed sample period that excludes the GFC (2009-06-01 to 2025-08-29). Our exclusion of the GFC is motivated and described under Section V.C.

A. Constructing Box Rates and Convenience Yields

The convenience yield of tenor n is estimated as the spread between the interest rate one can gain through a box-trade on European options on the SPX with expiry in n years – the "box rate" with an n-year tenor – and the yield on nominal Treasuries with an n-year tenor. Thus, to construct convenience yields, we must construct time series of box rates. To construct time series of box rates from European options on the SPX, we utilize the entire panel data set of SPX options described under Section IV. On each day, we group all contracts by their days until expiry, which we express as fractional years until expiry n. Within each pair, we follow the methodology of van Binsbergen et al. (2022). Matched pairs are created between each call and put of the same strike price, K, and the following regression is estimated:

$$C_{n,K} - P_{n,K} = \alpha + \beta_n \times K + e_{n,K} \tag{5}$$

If put-call parity holds, the coefficient, β_n from the regression is proportional to the continuously compounded annualised risk-free interest rate with a n-year tenor:

$$\beta_n = -e^{-r^{box}_{t,n} \cdot n}$$

$$r^{box}_{t,n} = -\frac{\ln\left[\frac{1}{\beta_n}\right]}{n}$$

On a given day, we perform this computation for each n to obtain pairwise observations of

the continuously compounded annualised risk-free box rate over different tenors:

$$(r_{t,n}^{box}, n)$$

With this, we apply NSS estimation (see Appendix A for details) of a complete box rate yield curve. From the complete yield curve, we can observe box rates at 1-, 2-, and 3-year tenors. We then calculate convenience yields at each tenor n:

$$r_{t,n}^{conv} = r_{t,n}^{box} - r_{t,n}^{TSY}$$

We perform this computation each day to compute daily time series of convenience yields at 1-, 2-, and 3-year tenors. As the raw convenience yields contain noise induced by estimation error in the NSS-yield curve, differences between mid prices and transaction prices, and slight violations of put-call parity, we apply the time series outlier filter of Hampel (1974)[5] and the time series smoothing filter of Kalman (2011) to stabilize the time series for each tenor. The estimated convenience yields for 1-, 2-, and 3-year tenors respectively are presented in Figure 4.

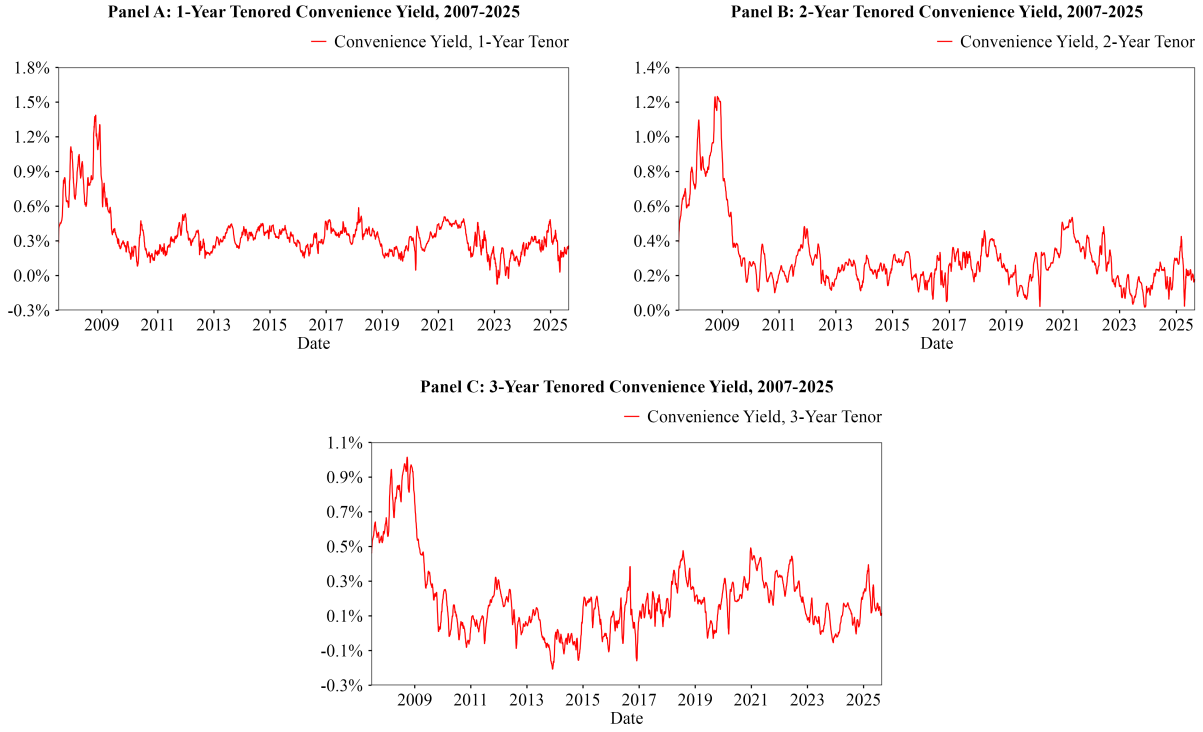


Figure 4. *Continuously Compounded Annual Convenience Yields, at 1-Year, 2-Year, and 3-Year Tenors*

B. Seasonal Adjustment

It is well documented that U.S. inflation exhibits substantial seasonality with one-year cycles (Wright, 2018) which translates into seasonality in prices and yields of inflation-indexed securities. In theory, this should not affect our mispricings A^1 and A^2, since any inflation-seasonality should affect both inflation-dependent legs of equations (1) and (2) – the inflation swap and

[5]Calibrated using window half-width of 5 and outlier threshold of 3.

the TIPS bond – equally, meaning no seasonality shows up in the mispricings A^1 and A^2. Yet, when observing the mispricings in Figure 5, a seasonal pattern is apparent. This is especially visible when inspecting the sub-sample which excludes the GFC.

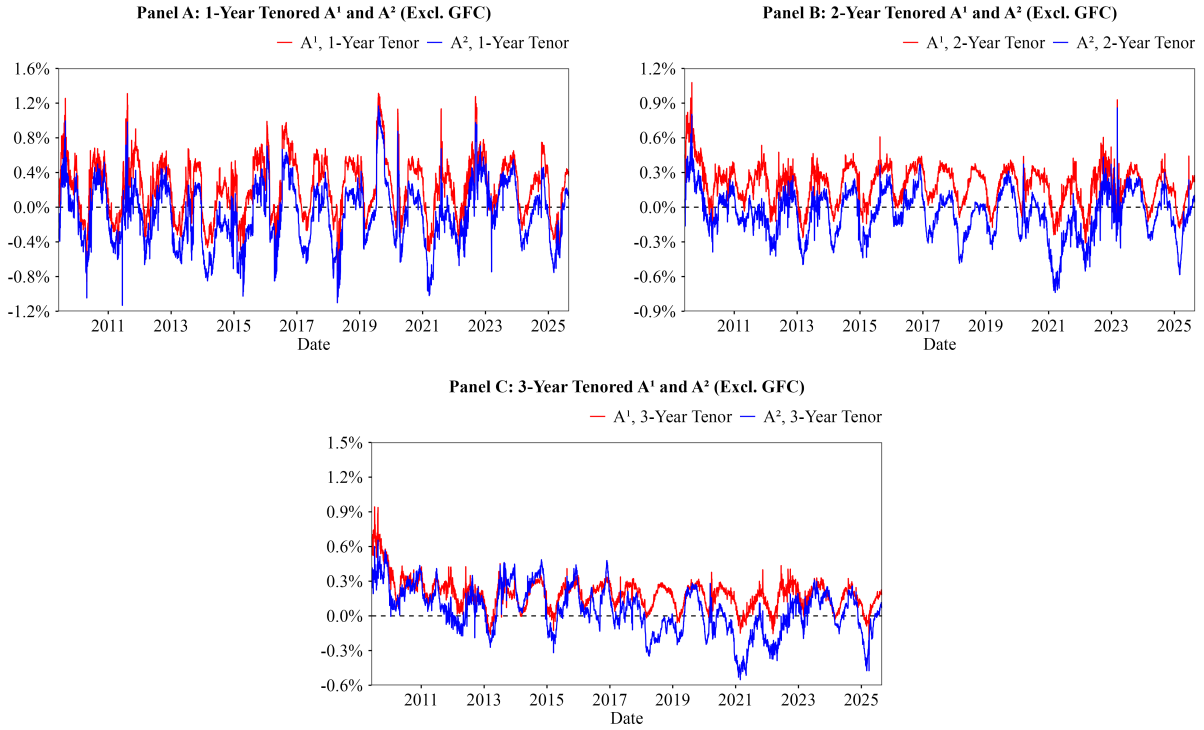


Figure 5. *Non-Seasonally Adjusted A^1 and A^2, at 1-Year, 2-Year, and 3-Year Tenors, Excluding the Great Financial Crisis*

The observed seasonality in the mispricings is an artefact of the NSS-estimation of TIPS yields. The NSS-parameters are estimated by fitting an NSS-model to observations of individual TIPS, which do not necessarily have integer-year tenors (Gürkaynak et al., 2010). Thus, the individual TIPS that underlie the NSS-estimation do not cover entire inflation-seasonality cycles. On the other hand, the inflation swap rates are based on actual observations of integer-year tenored inflation swaps, which do cover entire inflation-seasonality cycles. The result is a mismatch in the degree of seasonality that is priced by the inflation swap contra the TIPS bond. This means that the two inflation-indexed securities no longer cancel out their respective inflation-seasonality, ultimately producing seasonality in the mispricings. This difference in seasonality between inflation swaps and TIPS bonds is confirmed through direct observation in appendix B.

This is not an invalidation of the no-arbitrage condition. If we could observe true integer-year tenored TIPS yields, they would exhibit the same degree of seasonality as the inflation swaps. In this case, since the TIPS yields are estimated including TIPS with a non-standard number of days left until maturity, the estimated TIPS yields incorrectly carry a different degree of seasonality than the inflation swaps do.

To adjust for this, we de-seasonalise the estimated mispricing A^1. To preserve any seasonality that possibly exists in the convenience yield itself, we define A^2 as a function of the de-seasonalised mispricing A^1, adjusting for the convenience yield.

The de-seasonalising of A^1 is achieved by using the U.S. Census Bureau's X-13 ARIMA-SEATS (referred to as X-13) procedure (Findley, 2005; Findley, Monsell, Bell, Otto, and Bor-Chung Chen 1998) on each tenor (1-, 2-, and 3-year) separately. We choose this procedure over other simpler methods like regressions on monthly dummies or STL (Cleveland, Cleveland, McRae, and Terpenning, 1990) to allow for the seasonal pattern to change over time and to robustly account for large dislocations in the data. Extensive robustness checks of this procedure and its effect on our results are conducted under Section VII.D.

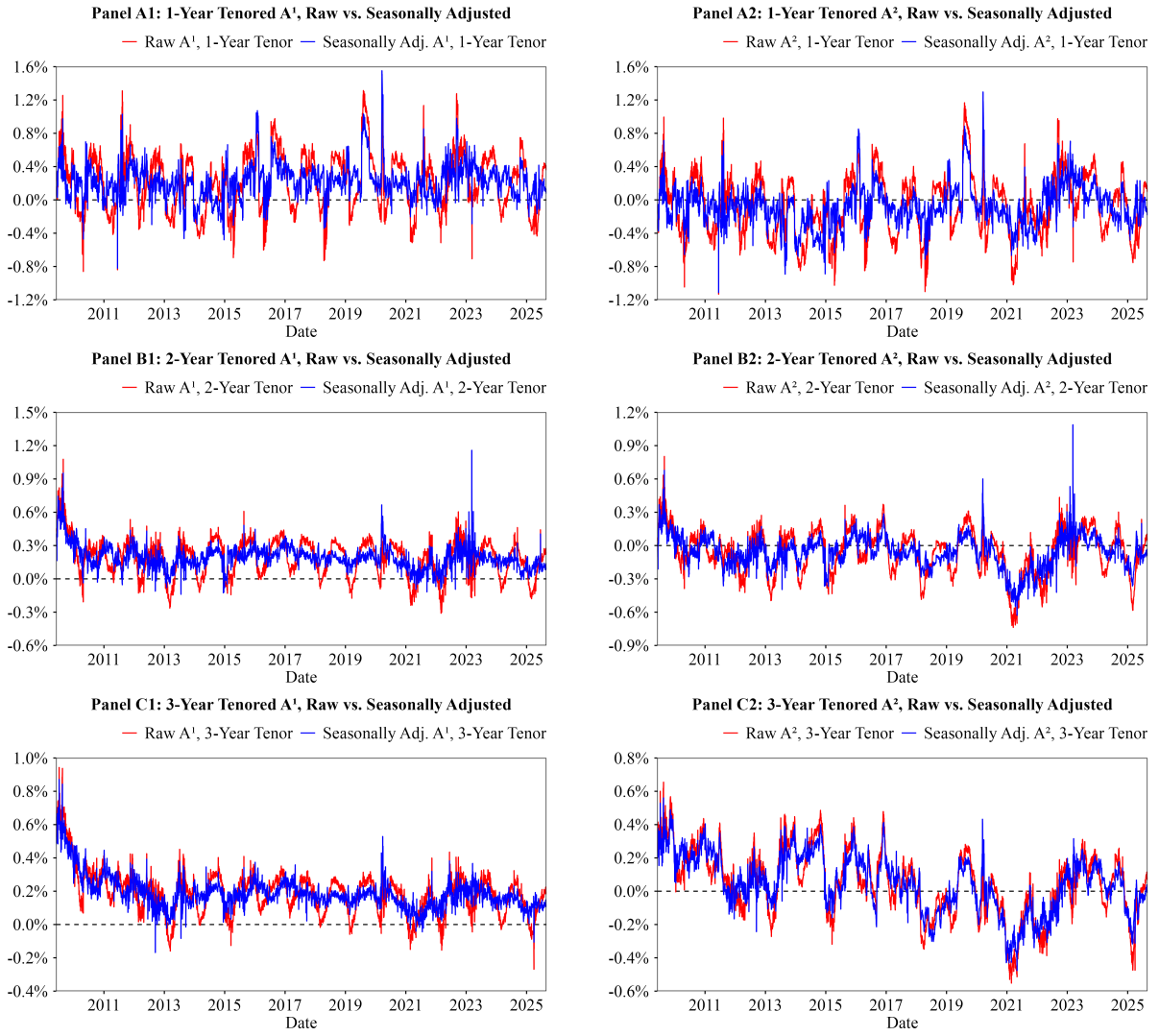


Figure 6. *Raw vs. Seasonally Adjusted A^1 (Left) and A^2 (Right), Excluding the Great Financial Crisis, at 1-Year (Top), 2-Year (Middle), and 3-Year (Bottom) Tenors*

C. Exclusion of the Great Financial Crisis

The convenience yield of nominal Treasury securities is documented to exhibit severely different behaviors during crisis episodes, compared to during non-crisis episodes (Krishnamurthy and Vissing-Jorgensen, 2012). It is likely that investors value the non-pecuniary conveniences of safety, liquidity and collateral higher during times of market stress. As such, it is reasonable to expect the economic significance of the convenience yield to be larger during major crises, such as the GFC. A question of interest is then whether the convenience yield is still econom-

ically significant in the context of the TIPS-Treasury bond puzzle during non-crisis episodes. To investigate this, we also execute all tests on a sub-sample which excludes the GFC.

Following the timeline laid out by Brunnermeier (2009), we treat August 2007 as the onset of the financial crisis and define the crisis period as August 2007 to June 2009, consistent with the NBER recession trough (NBER, 2010). All our main results are estimated and presented on both our full sample and our subsample excluding the GFC.

VI. Results

A. Research Question 1

There are two parts of this question that have been highlighted by researchers: (i) Market prices of TIPS and Treasury securities consistently deviate from no-arbitrage conditions, such that the mispricing defined in equation (4) is almost always non-zero, and (ii) these deviations are consistent in direction, such that the mispricing defined in equation (4) is consistently positive.

We find that adjusting for the convenience yield performs very well in resolving (ii), and is useful, although not exhaustive, in resolving (i). After adjusting for the convenience yield in accordance with equation (3), the average absolute level of deviations from no-arbitrage conditions is decreased. Still, deviations from zero-mispricing remain after accounting for the convenience yield. Adjusting for the convenience yield can thus decrease the average absolute level of arbitrage that exists in the TIPS-Treasury bond puzzle, but cannot remove all no-arbitrage violations entirely. While some absolute level of arbitrage remains, the convenience yield is able to bring the mean mispricing much closer to zero, such that the mispricing is no longer consistent in direction. While this is evidence that the convenience yield is of the right size and direction to account for nominal Treasury securities being systematically rich relative to TIPS-plus-inflation-swaps, it is not evidence that the convenience yield actually drives this mispricing.

Figure 7 presents the mispricings, before adjusting for convenience yield A^1 and after adjusting for convenience yield A^2 over our full sample, for 1-, 2-, and 3-year tenors respectively.

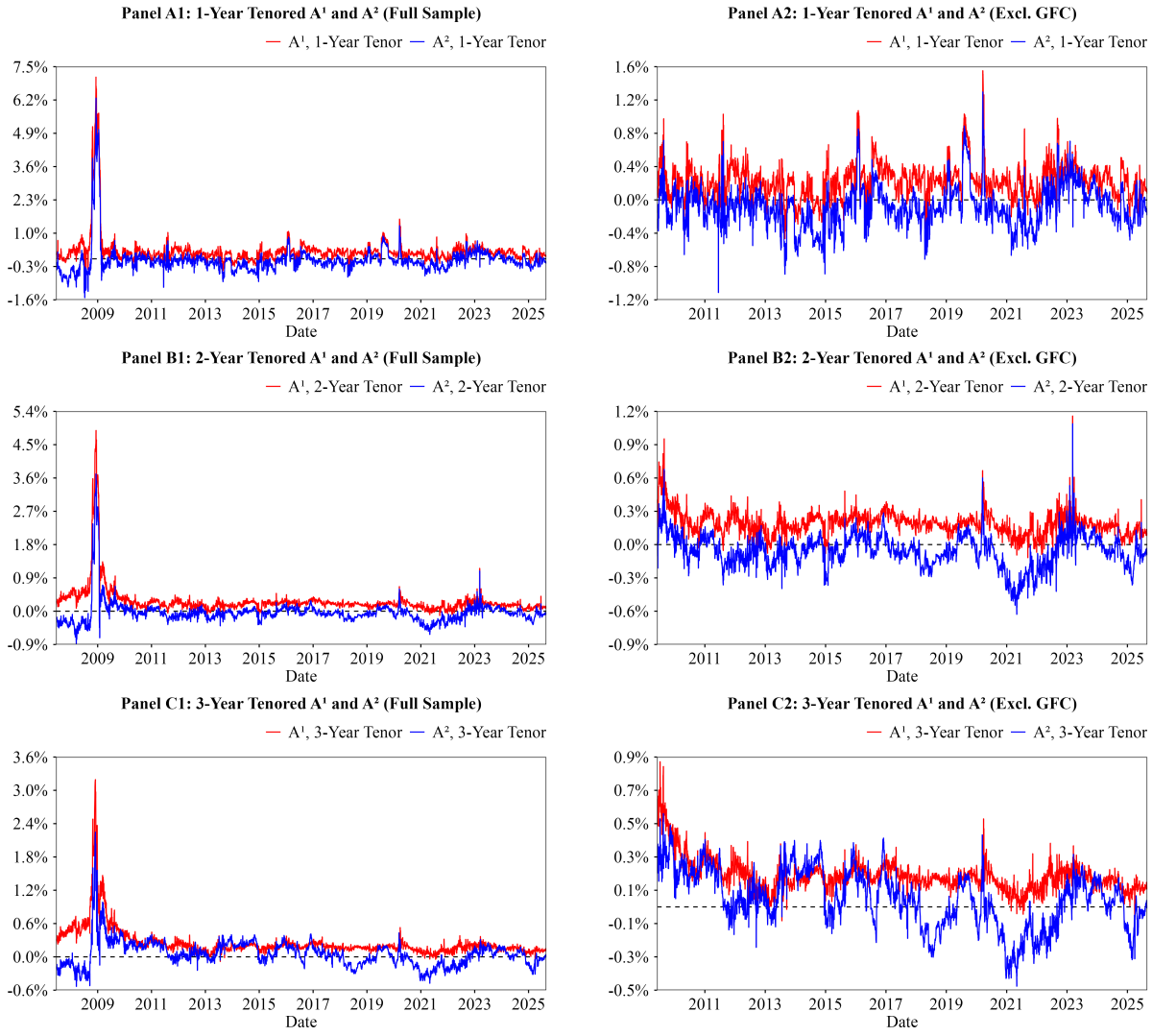


Figure 7. *Basic Mispricing (A^1) and Adjusted for Convenience Yields (A^2), Full Sample (Left) and Excluding the Great Financial Crisis (Right), at 1-Year (Top), 2-Year (Middle) and 3-Year (Bottom) Tenors*

Adjusting for the convenience yield appears to act as a level-shifter to the mispricing. The mispricing does not appear less volatile after adjusting for the convenience yield, but its oscillation appears more centered around zero.

Summary statistics of the average mispricing, before adjusting for convenience yield A^1 and after adjusting for convenience yield A^2, are presented in Table 4, for 1-, 2-, and 3-year tenors, quoted in percentage points. Table 5 reports the same statistics for the subsample excluding the GFC.

Table 4. *Summary Statistics of Inflation-Swap Arbitrage: Haubrich Relation vs. Convenience Yield Adjusted, at 1-Year, 2-Year, and 3-Year Tenors*

	1-Year Tenor		**2-Year Tenor**		**3-Year Tenor**	
Statistic	A^1	A^2	A^1	A^2	A^1	A^2
Mean	0.3059	−0.0456	0.2699	−0.0370	0.2533	0.0540
MAV	0.3327	0.2818	0.2714	0.1715	0.2536	0.1711
SD	0.6341	0.5788	0.4216	0.3435	0.2766	0.2377
Min	−0.8205	−1.5175	−0.1396	−0.8731	−0.1686	−0.5336
Max	7.5362	6.2745	4.8851	3.7092	3.1934	2.2465

Note: MAV denotes mean absolute value. SD denotes sample standard deviation.

Table 5. *Summary Statistics of Inflation-Swap Arbitrage: Haubrich Relation vs. Convenience Yield Adjusted, at 1-Year, 2-Year, and 3-Year Tenors, Excluding the Great Financial Crisis*

	1-Year Tenor		**2-Year Tenor**		**3-Year Tenor**	
Statistic	A^1	A^2	A^1	A^2	A^1	A^2
Mean	0.2218	−0.0764	0.1905	−0.0580	0.1890	0.0487
MAV	0.2476	0.1972	0.1921	0.1193	0.1893	0.1454
SD	0.2105	0.2470	0.1035	0.1446	0.0991	0.1707
Min	−0.8205	−1.1129	−0.1396	−0.6285	−0.1686	−0.4766
Max	1.5517	1.2967	1.1584	1.0884	0.8717	0.5603

Note: MAV denotes mean absolute value. SD denotes sample standard deviation.

Throughout our full sample period, adjusting for convenience yield brings the average mispricing from 31 basis points to −5 basis points per annum over a 1-year tenor, from 27 to −4 basis points per annum over a 2-year tenor, and from 25 to 5 basis points per annum over a 3-year tenor. This is expected, as the convenience yield is circa 30 basis points per annum, and larger at short tenors. Interestingly, the mispricing is not consistently positive on average after adjusting for the convenience yield, rather the average mispricing is negative over 1-year and 2-year tenors.

Beyond bringing the average mispricing closer to zero, adjusting for convenience yield brings the average absolute deviation from no-arbitrage conditions, at which mispricing should be zero, down from 33 basis points per annum to 28 basis points per annum over a 1-year tenor, from 27 to 17 basis points per annum over a 2-year tenor, and from 25 to 17 basis points per annum over a 3-year tenor. This supports the hypothesis that adjusting for convenience yields improves the adherence to no-arbitrage conditions. The results hold with similar magnitudes in the subsample excluding the GFC.

Formal statistical tests on whether accounting for the convenience yields reduces the mispricing (equation 4) are presented in Table 6 for the full sample, and Table 7 for the subsample excluding the GFC.

Table 6. *Newey-West Mean Difference Test of Absolute Inflation Swap Rate Arbitrage, Full Sample*

Term	**1-Year**	**2-Year**	**3-Year**
Mean Difference	0.0509***	0.0999***	0.0825***
	(0.0190)	(0.0144)	(0.0135)
N	4547	4547	4547

Table 7. *Newey-West Mean Difference Test of Absolute Inflation Swap Rate Arbitrage, Excluding the Great Financial Crisis*

Term	**1-Year**	**2-Year**	**3-Year**
Mean Difference	0.0505***	0.0729***	0.0439***
	(0.0137)	(0.0094)	(0.0084)
N	4058	4058	4058

Accounting for convenience yield provides a statistically significant decrease in the average absolute deviation from no-arbitrage conditions, at which the mispricing should be zero, at the 1-year tenor at the 1% significance level, and at 2-year and 3-year tenors at the 0.1% significance level in the full sample. The results hold with similar magnitude and significance in the sub-sample excluding the GFC.

B. Research Question 2

We find that the convenience yield is a predictor that can explain up to 35 percent of the variation in the mispricing between TIPS, nominal Treasury securities, and inflation swaps over our full sample with slope estimates that align well in both direction and magnitude to the economic expectation. However, this effect is almost entirely carried by the GFC. Despite the GFC generating much of the in-sample fit, its high autocorrelation means it contributes far less information than its raw observation count suggests. Thus, we see that the effect is not statistically significant at 1- and 2-year tenors despite the high R^2. When excluding the GFC, we find negative slope estimates and R^2 values below 0.03, with our slope estimate being statistically insignificant at 2- and 3-year tenors.

Together with the results from Section VI.A, a reasonable interpretation is that the convenience yield does act as a constant shifter in the no-arbitrage relationship, but that the convenience yield does not vary in an economically meaningful way outside of severe crisis episodes. If economic variations of the convenience yield are small during "normal times", noise introduced by the box-rate estimation is likely to dominate the convenience yield signal, making it a poor statistical predictor of the mispricing. A more skeptical interpretation is that the convenience yield is not a significant factor driving the TIPS-Treasury bond puzzle, invalidating the hypothesis laid out by Fleckenstein et al (2014), but that since the convenience yield is positive on average and the basic mispricing A^1 is too, adjusting for the convenience yield still appears to reduce the arbitrage.

Figure 8 shows A^1 and the convenience yield in the full sample and the subsample excluding the GFC illustrating how they co-move.

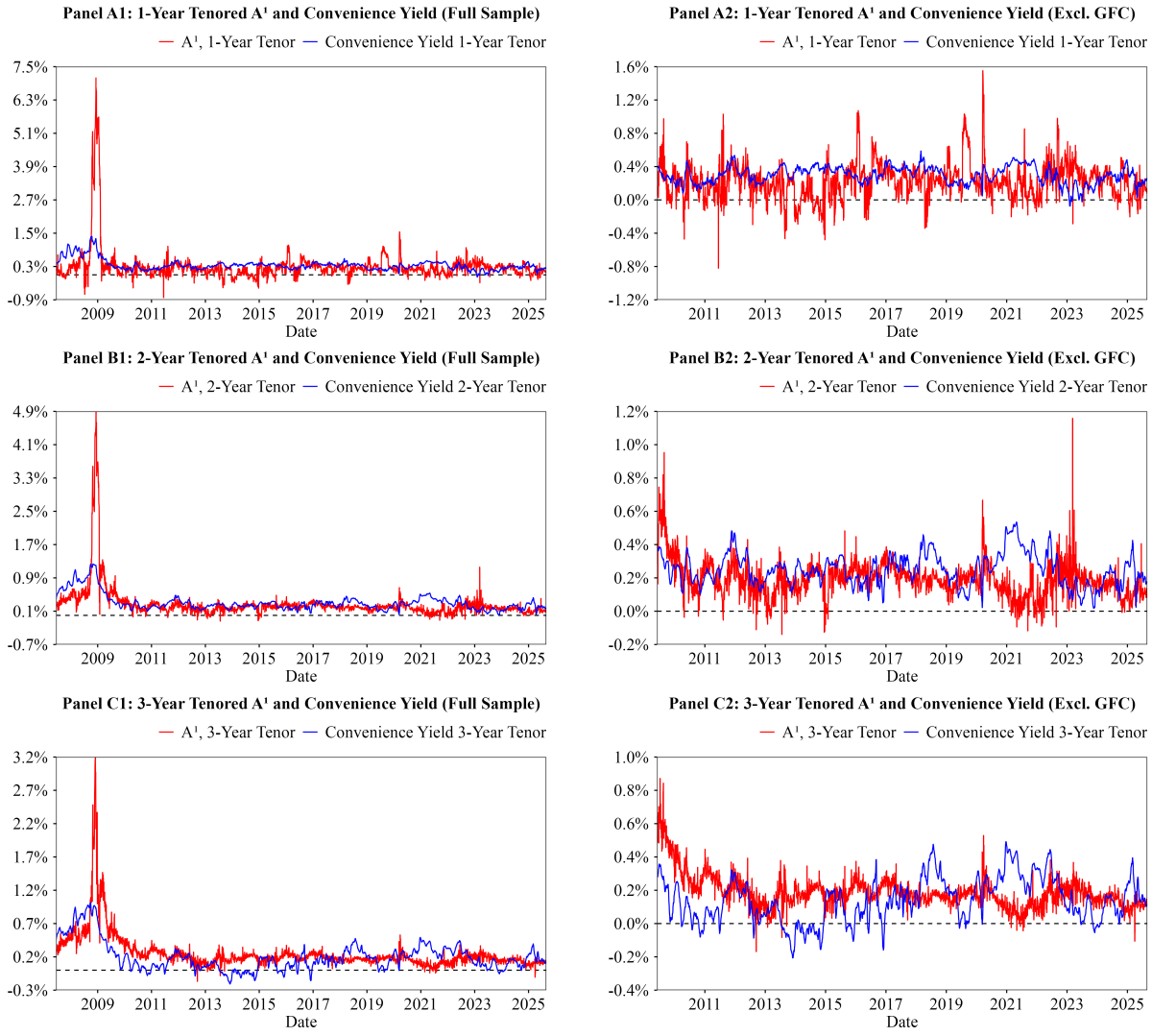


Figure 8. *Mispricing (A^1) and Convenience Yields, Full Sample (Left) and Excluding the Great Financial Crisis (Right), at 1-Year (Top), 2-Year (Middle), and 3-Year (Bottom) Tenors*

The variables appear correlated during the GFC, but the same correlation cannot be seen in periods after the GFC. We also note that the convenience yield appears to be a leading indicator of the unadjusted mispricing A^1 during the GFC. We investigate this effect further in Section VII.G.

We test whether the co-movement between the mispricing A^1 and the convenience yield of nominal Treasury Securities is statistically significant by estimating the following regression:

$$A^1_{t,n} = \alpha + \beta \times r^{conv}_{t,n} + e_{t,n}$$

Regression results for the 1-year, 2-year, and 3-year tenors are presented for the full sample in table 8, and excluding the GFC in Table 9.

Table 8. *Regressing Inflation Swap Mispricing on Convenience Yield, Over 1-Year, 2-Year, and 3-Year Tenors, Full Sample*

Term	**1-Year**	**2-Year**	**3-Year**
Intercept	−0.1712	−0.1041	0.1122**
	(0.6513)	(0.3885)	(0.0440)
Slope	1.3575	1.2185	0.7082**
	(2.2112)	(1.5451)	(0.3220)
R-squared	0.1791	0.3474	0.3149
N	4547	4547	4547

Note: Newey-West standard errors are used due to serial correlation of the timeseries.

Table 9. *Regressing Inflation Swap Mispricing on Convenience Yield, Over 1-Year, 2-Year, and 3-Year Tenors, Excluding the Great Financial Crisis*

Term	**1-Year**	**2-Year**	**3-Year**
Intercept	0.3108***	0.2047***	0.1973***
	(0.0360)	(0.0143)	(0.0070)
Slope	−0.2982***	−0.0573	−0.0590
	(0.1131)	(0.0657)	(0.0549)
R-squared	0.0210	0.0028	0.0061
N	4058	4058	4058

Note: Newey-West standard errors are used due to serial correlation of the timeseries.

In the full sample, the intercept is statistically indistinguishable from zero over 1- and 2-year tenors, indicating no residual mispricing once the convenience yield is accounted for. This is consistent with the finding that adjusting for the convenience yield can bring the average arbitrage closer to zero.

The convenience yield can explain 18, 35, and 31 percent of the mispricing at 1-year, 2-year, and 3-year tenors respectively in the full sample. The lower explanatory power at the 1-year tenor may be an artefact of the higher uncertainty in the TIPS extrapolation to the 1-year tenor as explained in Section IV. The slopes of the convenience yield on the mispricing A^1 are not statistically significant across the 1-year and 2-year tenors, but the magnitude in slope coefficients over tenors reveals two interesting phenomena. A 1 percentage point increase in convenience yield should lead to a 1 percentage point increase in the mispricing A^1 over all tenors, i.e. the slope coefficients should be strictly 1 – if convenience yields are exogenous from other predictors of mispricing. A possible explanation for the fact that the slope estimate is larger than 1 for the 1-, and 2-year tenors is that the convenience yield co-moves with other predictors of mispricing, such as a liquidity premium in TIPS. This makes economic sense: liquidity premia and the convenience yield, which are both "flight-to-safety" premia, are higher in times of high macroeconomic stress. Indeed, D'Amico et al. (2018) document a peak in their estimated TIPS liquidity premium around the GFC, during which we and van Binsbergen et al. (2022) also document a peak in convenience yields. An explanation of the observed decrease in slope coefficient over tenors is that random measurement error in the regressor – in this case noise in convenience yields induced by NSS-estimation – grows larger with longer tenors, causing greater attenuation of the estimated effect.

The same results do not hold when excluding the GFC. The intercept is statistically signif-

icantly positive across all tenors, suggesting that an average mispricing of 20-31 basis points remains even when the convenience yield is zero. Further, the estimated coefficients of the convenience yield on the mispricing A^1 are negative, contradicting the economic expectation outlined in Section III.B. This finding is statistically significant at the 1% level at the 1-year tenor. An economically sensible interpretation is that, outside of crisis episodes, the convenience yield negatively co-moves with a positive unobservable predictor or positively co-moves with a negative unobservable predictor of the mispricing. Together with the full-sample results, an interpretation is that the convenience yield co-moves with other market-stress factors, for example a TIPS liquidity premium, during crisis episodes, but does not co-move or co-moves negatively with the same factors during non-crisis episodes. These interpretations should be taken with caution, however, as they are largely based on statistically insignificant estimates of slope coefficients.

VII. Diagnostics and Robustness Checks

A. Comparing Box Rates Across Underlying Indices

Following van Binsbergen et al. (2022), we have chosen to use the SPX as the underlying index for the options contracts used to calculate the box rate. However, it is worth acknowledging that this is not an insignificant choice. Figure 9 presents time series of box rates, continuously compounded to annual rates, with a 7-day tenor (the most actively traded maturities), using SPX and RUT separately.

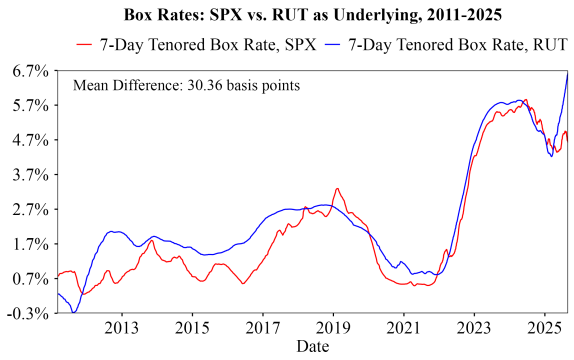


Figure 9. *7-Day Tenored Box Rates Over Time: SPX vs. RUT as Underlying Index*

We find that there is a significant and persistent difference in levels between box rates from SPX and RUT, such that box rates using RUT are on average 30 basis points higher than box rates using SPX. The choice of index may thus significantly alter our results. We choose to focus this paper on box rates implied from the SPX options as these are the most actively traded option contracts in the world (Schwartz, 2026), decreasing the risk of liquidity premiums and put-call parity violations, and acknowledge this finding as a limitation of the study.

This observation of a persistent spread between box rates of different indices creates substantial questions for future research: How can these spreads be explained? Does there exist an arbitrage opportunity by going long a box trade on one underlying while going short a box trade on another underlying? Would such arbitrage opportunities remain after accounting for

bid-ask spreads and transaction costs?

B. The Accuracy of NSS-Estimations

One concern is that NSS-estimation does not provide an adequate estimate of box rate yield curves. As a robustness check of the NSS-estimation procedure, we can compare the yield-curve-estimated convenience yields to actual observable convenience yields. For this, we create subsets of the options data to only contain contracts with exactly 1 year, 2 years, and 3 years left until expiry respectively, and execute the calculations necessary for equation (5) with $n = 1$, $n = 2$, and $n = 3$ respectively, on a daily basis. Observable convenience yields are compared to yield-curve-estimated convenience yields in Figure 10.

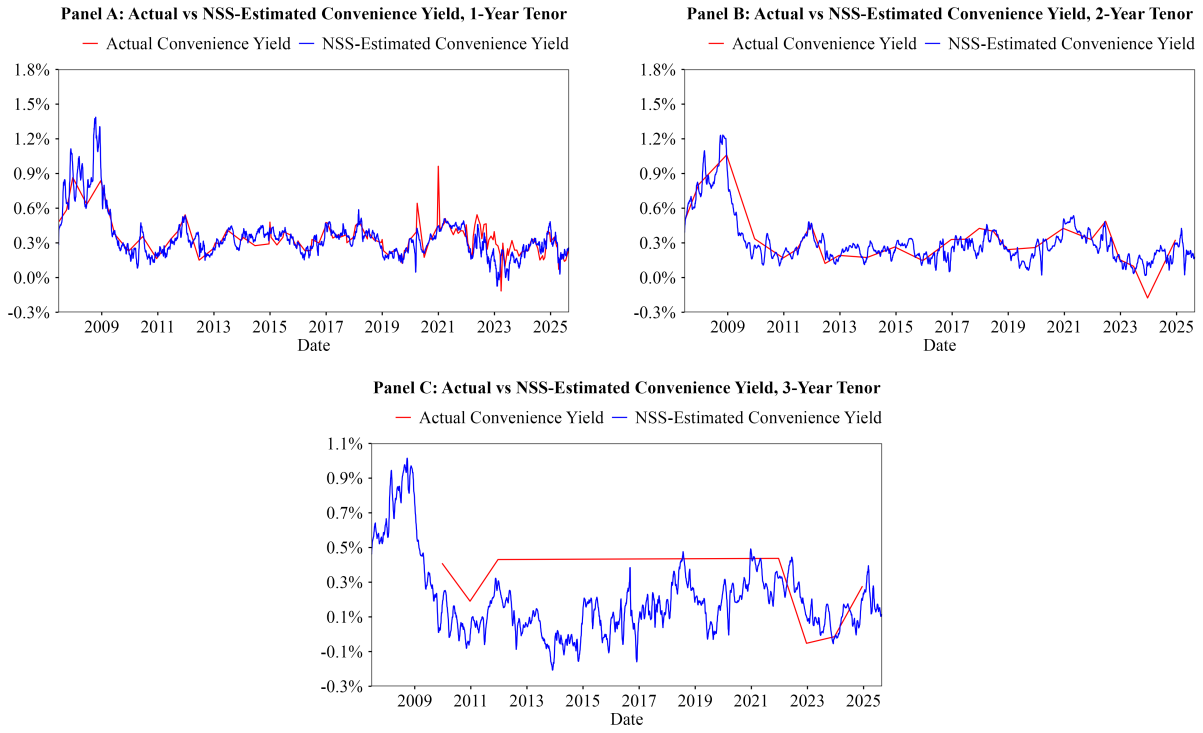


Figure 10. *Observable Convenience Yields vs. NSS-Estimated Convenience Yields, at the 1-Year, 2-Year, and 3-Year Tenors*

Note: The actual convenience yields are observed at low frequencies, and linear interpolation is used to create continuous time series.

As can be observed, the frequency at which 2-year and 3-year tenored SPX options are outstanding is very low, which is why NSS-estimation is beneficial. It can be observed that the NSS-estimated convenience yield provides a good estimation of the actual convenience yield at the 1-year and 2-year tenors, with the benefit of a higher frequency of observations but with the drawback of estimation error around 2007-2008 for the 1-year tenor, and around 2025 for the 2-year tenor. The fit of the NSS-estimated convenience yield is hard to assess for the 3-year tenor, as there are only 7 days in our data set during which an options contract with 3 years left until expiry is observed.

C. Controlling Put-Call Parity Adherence

A further concern is that, if put-call parity is violated, the equations on which our estimation of the box rate relies no longer hold true. This is especially problematic if the extent to which put-call parity holds varies depending on the tenor of the options that are traded, such that shorter-dated options have a different degree of bias compared to longer-dated options, as this could induce severe bias in the full NSS-estimated yield curve. As noted by van Binsbergen et al. (2022), the R^2 obtained when estimating regression (6) is an adequate measure of the extent to which put-call parity holds in the observed cross-section. Thus, we can execute the following checks of put-call parity adherence: (i) Plot the observed R^2 over time, to understand the time-varying structure of put-call parity adherence, and (ii) compare R^2 of short-tenored options to the R^2 of long-tenored options over time, to understand whether a significant difference exists, and if so how it evolves. This is presented in Figure 11.

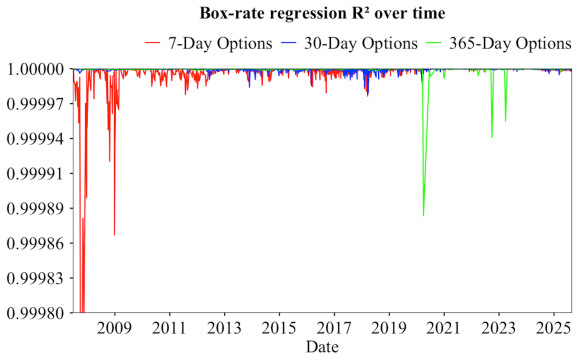


Figure 11. *R^2 of The Put-Call Parity Regression, For 7-Day, 30-Day, and 365-Day Options*

Note: The Y-Axis is trimmed for better visualization; The lowest R^2 was measured as 0.9992 for 7-day options on 2007-10-26

The R^2 never drops below 0.999 across all tested tenors throughout our entire sample period, implying that any concern of put-call parity being violated is negligible. Still, there is a slight bias towards lower R^2 in short-dated options, which is explained when looking at Tables 1-3; short-dated options have larger bid-ask spreads, which means that using mid-prices as acceptable proxies for transaction prices is a stronger assumption, and thus may introduce larger noise.

D. Validity of The Seasonality Adjustment

As a diagnostic of the seasonality adjustment, we re-create our main results without the seasonal adjustment applied. If the seasonal adjustment is valid, the estimated coefficients should be unchanged, and the only benefit should be the reduction of unrelated noise which otherwise would understate the statistical significance of our results.

First, we test whether the absolute value of A^1 is on average larger than the absolute value of A^2 without the seasonal adjustment. Results are presented for the full sample in Table 10, and excluding the GFC in Table 11.

Table 10. *Newey-West Mean Difference Test of Absolute Inflation Swap Rate Arbitrage, Non-Seasonally Adjusted, Full Sample*

Term	**1-Year**	**2-Year**	**3-Year**
Mean Difference	0.0367*	0.0842***	0.0718***
	(0.0281)	(0.0191)	(0.0199)
N	4547	4547	4547

Table 11. *Newey-West Mean Difference Test of Absolute Inflation Swap Rate Arbitrage, Non-Seasonally Adjusted, Excluding the Great Financial Crisis*

Term	**1-Year**	**2-Year**	**3-Year**
Mean Difference	0.0350**	0.0515***	0.0296**
	(0.0200)	(0.0175)	(0.0132)
N	4058	4058	4058

The results observed in Section VI.A hold with the same sign and similar magnitude without the seasonal adjustment, with the drawback of larger standard errors induced by the variation in the dependent variable that stems from inflation-seasonality.

Further, we test whether the regression results of Section VI.B hold without the seasonal adjustment. Results are presented in Table 12 for the full sample and Table 13 for the sample excluding the GFC.

Table 12. *Regressing Mispricing (Non-Seasonally Adjusted) on Convenience Yield, Over 1-Year, 2-Year, and 3-Year Tenors, Full Sample*

Term	**1-Year**	**2-Year**	**3-Year**
Intercept	−0.1956	−0.0991	0.1163**
	(0.6745)	(0.4169)	(0.0523)
Slope	1.4286	1.2040	0.6893*
	(2.3044)	(1.6637)	(0.3607)
R-squared	0.1639	0.3069	0.2726
N	4547	4547	4547

Table 13. *Regressing Mispricing (Non-Seasonally Adjusted) on Convenience Yield, Over 1-Year, 2-Year, and 3-Year Tenors, Excluding the Great Financial Crisis*

Term	**1-Year**	**2-Year**	**3-Year**
Intercept	0.2869***	0.2362***	0.2093***
	(0.1013)	(0.0303)	(0.0116)
Slope	−0.2155	−0.1806	−0.1413
	(0.2999)	(0.1382)	(0.0951)
R-squared	0.0042	0.0107	0.0198
N	4058	4058	4058

The full-sample results are robust to the seasonality-adjustment, with estimated slopes closely aligned in magnitude and insignificance to the main empirical results, R^2 that is slightly lower across all tenors – since some variation in the dependent variable is now purely driven by the inflation seasonality artefact –, and intercepts that remain insignificant across the 1-year

and 2-year tenors. The results for the sub-sample period excluding the GFC also remain stable without the seasonality-adjustment: intercepts are statistically significant at 20-30 basis points per annum, and estimated coefficients of the convenience yield on the mispricing are negative but insignificant.

E. Diagnostics of the Newey-West Test of Mean Difference

The main concerns regarding the validity of the Newey-West test of mean difference presented in Section VI.A are sensitivity to the automatic bandwidth selection and sensitivity to outlier observations. In Table 14, we present tests of the sensitivity of the significance of our results to the bandwidth selection by inspecting p-values of static lags at 1-year, 2-year, and 3-year tenors respectively.

Table 14. *One-Sided Newey West P-Values For $E[|A^1|] > E[|A^2|]$, With Lags From 21 to 1260, at 1-Year, 2-Year, and 3-Year Tenors*

Method	**1-Year Mean Diff**	**2-Year Mean Diff**	**3-Year Mean Diff**
Manual L=21	0.0509***	0.0999***	0.0825***
	(0.0195)	(0.0152)	(0.0146)
Manual L=63	0.0509**	0.0999***	0.0825***
	(0.0261)	(0.0223)	(0.0202)
Manual L=126	0.0509**	0.0999***	0.0825***
	(0.0301)	(0.0285)	(0.0259)
Manual L=252	0.0509*	0.0999***	0.0825***
	(0.0341)	(0.0349)	(0.0324)
Manual L=504	0.0509*	0.0999***	0.0825**
	(0.0316)	(0.0393)	(0.0388)
Manual L=756	0.0509**	0.0999***	0.0825**
	(0.0268)	(0.0404)	(0.0410)
Manual L=1260	0.0509***	0.0999***	0.0825**
	(0.0218)	(0.0413)	(0.0437)
Newey-West (1994)	0.0509***	0.0999***	0.0825***
	(0.0155)	(0.0113)	(0.0126)

In the full sample, the convenience yield's ability to reduce no-arbitrage violations are significant at the 1% level across all static lags tested for the 2-year tenor, at the 5% level at all static lags tested for the 3-year tenor, and at the 5% level at all static lags tested except for two for the 1-year tenor. The results of Section VI.A appear robust to different lag selections.

In Table 15, we test the sensitivity of the estimated mean difference to outlier observations, by (i) excluding observations outside of the 99th and 95th percentiles respectively, and (ii) winsorizing observations to the 99th and 95th percentiles respectively.

Table 15. *Newey-West Test of Mean Difference for $E[|A^1|] > E[|A^2|]$ Excluding Outliers and Winsorizing at the 99th and 95th Percentile, at 1-Year, 2-Year, and 3-Year Tenors*

Specification	**1-Year Mean Diff**	**2-Year Mean Diff**	**3-Year Mean Diff**	N
Baseline	0.0509*	0.0999***	0.0825**	4547
	(0.0341)	(0.0393)	(0.0410)	
Drop >99th pct	0.0412*	0.0902***	0.0742**	4501
	(0.0305)	(0.0335)	(0.0349)	
Drop >95th pct	0.0221	0.0699***	0.0557***	4319
	(0.0303)	(0.0258)	(0.0230)	
Winsorize at 99th	0.0513*	0.1003***	0.0793**	4547
	(0.0344)	(0.0395)	(0.0382)	
Winsorize at 95th	0.0314	0.0800***	0.0687**	4547
	(0.0269)	(0.0268)	(0.0296)	

The convenience yield's ability to reduce no-arbitrage violations remains significant at the 1% level for the 2-year tenor after winsorizing and observation-dropping at the 99th and 95th percentiles respectively. The effect remains significant at the 5% level for the 3-year tenor, but does not remain robustly significant at the 1-year tenor. The results of Section VI.A appear robust to the exclusion of outliers for the 2-year and 3-year tenors, but this robustness is weaker for the 1-year tenor.

F. Diagnostics of the OLS Regression

The main concerns regarding the validity of the OLS regression results are the sensitivity to the dynamic bandwidth selection in dealing with autocorrelated errors and the possible existence of unit roots causing spurious regression issues. Table 16 presents the full-sample slope coefficient p-values using a range of static lags.

Table 16. *Newey-West P-Values for the Slope Coefficient, With Lags from 21 to 1260 at 1-Year, 2-Year, and 3-Year Tenors*

Method	**1-Year Slope**	**2-Year Slope**	**3-Year Slope**
Manual L=21	1.3575**	1.2185***	0.7082***
	(0.5597)	(0.3538)	(0.1629)
Manual L=63	1.3575*	1.2185**	0.7082***
	(0.8133)	(0.5061)	(0.2374)
Manual L=126	1.3575	1.2185**	0.7082***
	(0.8801)	(0.5463)	(0.2727)
Manual L=252	1.3575	1.2185**	0.7082**
	(0.8578)	(0.5178)	(0.2813)
Manual L=504	1.3575**	1.2185***	0.7082***
	(0.6897)	(0.4186)	(0.2651)
Manual L=756	1.3575**	1.2185***	0.7082***
	(0.6137)	(0.3734)	(0.2517)
Manual L=1260	1.3575**	1.2185***	0.7082***
	(0.5538)	(0.3379)	(0.2476)
Newey-West (1994)	1.3575*	1.2185***	0.7082**
	(0.7403)	(0.4255)	(0.2958)

In the full sample, the convenience yield is a statistically significant predictor of the mispricing A^1 at 2-year and 3-year tenors at the 5% significance level or lower across all static

lags. The predictor fluctuates in statistical significance across the 1-year tenor, with the highest p-values being registered at a lag length of 1-2 years. A possible explanation for the statistical significance when lags are static is that it overweights the events of the GFC, during which our variables were exhibiting extraordinarily long and strong autocorrelation. A static lag designed to fit the full sample is likely too weak to handle the autocorrelation during the GFC, overstating the significance of the full-sample estimates. In Table 17, we present the same statistics for the sub-sample excluding the GFC and show that the sub-sample results presented in Section VI.B remain robust to lag selection.

Table 17. *Newey-West P-Values for the Slope Coefficient, With Lags from 21 to 1260 at 1-Year, 2-Year, and 3-Year Tenors, Excluding the Great Financial Crisis.*

Method	**1-Year Slope**	**2-Year Slope**	**3-Year Slope**
Manual L=21	−0.2982***	−0.0573	−0.0590
	(0.1091)	(0.0676)	(0.0517)
Manual L=63	−0.2982**	−0.0573	−0.0590
	(0.1472)	(0.0999)	(0.0762)
Manual L=126	−0.2982*	−0.0573	−0.0590
	(0.1674)	(0.1210)	(0.0900)
Manual L=252	−0.2982	−0.0573	−0.0590
	(0.1944)	(0.1419)	(0.1030)
Manual L=504	−0.2982	−0.0573	−0.0590
	(0.2206)	(0.1605)	(0.1077)
Manual L=756	−0.2982	−0.0573	−0.0590
	(0.2186)	(0.1696)	(0.1058)
Manual L=1260	−0.2982	−0.0573	−0.0590
	(0.1846)	(0.1705)	(0.1112)
Newey-West (1994)	−0.2982**	−0.0573	−0.0590
	(0.1463)	(0.1024)	(0.0895)

Further, we test for the existence of unit roots or stationarity of the convenience yield and the mispricing A^1 in Table 18, using ADF, PP and KPSS tests, in the full sample.

Table 18. *Unit Root and Stationarity Tests of the Convenience Yield and the Mispricing A^1, at 1-Year, 2-Year, and 3-Year Tenors, Full Sample. ADF H_0: Unit Root. PP H_0: Unit Root. KPSS H_0: Stationarity.*

Variable	**Tenor**	**ADF stat**	**ADF p**	**PP stat**	**PP p**	**KPSS stat**	**KPSS p**	**N**
A^1 (mispricing)	1-Year	−5.286	0.0100	−74.354	0.0100	1.607	0.0100	4547
Convenience yield	1-Year	−4.295	0.0100	−33.533	0.0100	9.093	0.0100	4547
A^1 (mispricing)	2-Year	−4.673	0.0100	−59.528	0.0100	5.374	0.0100	4547
Convenience yield	2-Year	−3.238	0.0813	−18.431	0.0953	9.732	0.0100	4547
A^1 (mispricing)	3-Year	−4.056	0.0100	−73.827	0.0100	11.133	0.0100	4547
Convenience yield	3-Year	−3.104	0.1109	−16.395	0.1955	5.505	0.0100	4547

Note: p-values for ADF and PP are interpolated and capped at 0.01 / 0.99.

The tests conflict in a diagnostic pattern: ADF and KPSS jointly reject for most series, which is logically inconsistent. One explanation is stationarity around structural breaks. ADF has low power against break alternatives (Perron, 1989) and KPSS is upward-biased by mean shifts. The 2-year and 3-year convenience yields are exceptions, with ADF marginal or failing to reject and PP failing to reject at conventional levels, suggesting near-unit-root behavior.

The relevant question for our main specification, however, is not whether the individual series are stationary but whether their linear combination is. If A^1 and the convenience yield are cointegrated, the residuals from the levels regression are stationary, OLS in levels is super-consistent (Stock, 1987), and the spurious-regression concern raised by Table 18 does not apply. Table 19 reports ADF and KPSS tests on the residuals from the main regression of Section VI.B.

Table 19. *Unit Root and Stationarity Tests of the Main Regression Residuals, at 1-Year, 2-Year, and 3-Year Tenors*

Tenor	Levels: ADF stat	Levels: ADF p	Levels: KPSS stat	Levels: KPSS p
1-Year	−5.928	0.0100	0.306	0.1000
2-Year	−5.503	0.0100	0.596	0.0230
3-Year	−4.970	0.0100	8.200	0.0100

ADF strongly rejects the unit root null at all tenors, providing Engle-Granger evidence of cointegration. KPSS agrees at the 1-year, is borderline at the 2-year, and rejects strongly at the 3-year tenor. The 3-year tenor KPSS rejection mirrors the diagnostic conflict in Table 18 and admits the same explanation: structural breaks induce mean shifts in the residuals that bias KPSS upward, even when the residuals are stationary within regimes. We interpret the evidence as cointegration around structural breaks (Gregory and Hansen, 1996), likely around crisis episodes, and proceed with the levels specification reading the slope as the average long-run equilibrium across regimes.

G. Lagged Effect of Convenience Yields on the Mispricing

We concretize the lagged structure of the convenience yield on the mispricing A^1 that was initially observed in Figure 8 through a distributed lag model with monthly lags of the convenience yield, up to 12 months, to predict the mispricing A^1. The estimated coefficients of the convenience yield on A^1 are presented in Figure 12 for the 1-year, 2-year, and 3-year tenors, along with 95% confidence intervals, over our full sample.

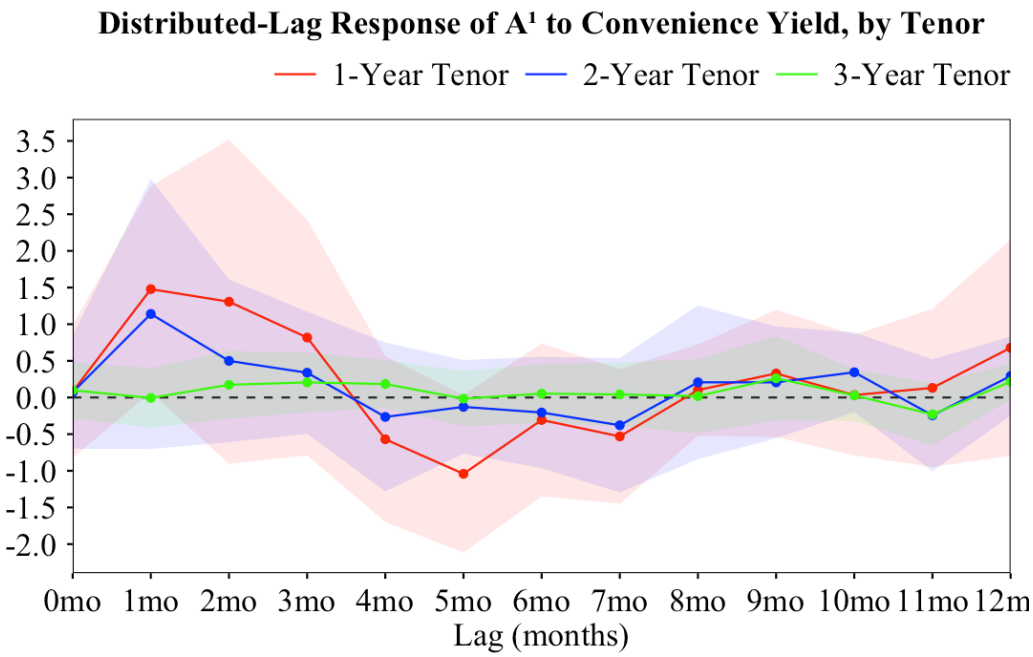


Figure 12. *Distributed-Lag Response of A^1 to the Convenience Yield, at 1-Year, 2-Year, and 3-Year Tenors, with Lags From 0 to 12 Months*

The convenience yield exhibits a lagged correlation to the mispricing A^1, such that full changes in the mispricing materialise up to three months after a positive movement in the convenience yield. However, these results are not statistically significant. This effect is stronger over shorter tenors, and statistically unobservable at the 3-year tenor.

VIII. Discussion

We empirically test an explanation for the TIPS-Treasury bond puzzle that has been suggested in the literature but not yet rigorously studied: that the convenience yield of nominal Treasury securities accounts for the observed deviations from no-arbitrage conditions between TIPS, nominal Treasuries, and inflation swaps. The average raw mispricing in our sample is between 25 and 31 basis points per annum, with an average absolute mispricing between 25 and 33 basis points. After adjusting for the convenience yield using box rates (van Binsbergen et al., 2022), the average mispricing falls to between −5 and +5 basis points, and the average absolute mispricing falls to between 17 and 28 basis points. The directional bias previously documented in the literature – where nominal Treasury securities are systematically expensive relative to TIPS-plus-inflation-swaps – disappears. However, from our regression of the mispricing on convenience yields, we find that variation in the convenience yield cannot consistently explain variation in the mispricing.

A. Interpretation of Results

Our headline finding is best framed as a directional result, not a variation result. Two distinct empirical exercises in the paper speak to different aspects of the convenience yield channel, and they yield different conclusions.

The first exercise, in Section VI.A, asks whether the average convenience yield is of the right size to account for the average mispricing. The answer is yes. The average convenience yield in our sample is approximately the same magnitude as the average puzzle, and adjusting for it brings the mean mispricing close to zero across all tenors, while reducing the average absolute deviation from no-arbitrage conditions by 5 to 10 basis points. This is the level result. We emphasize that this is a magnitude-consistency check rather than a causal identification: by construction, $A^2 = A^1 - r^{conv}$, so subtracting a positive quantity from a positively biased series will always pull the series toward zero. The non-trivial finding is that the magnitudes line up: the average convenience yield is neither far too small nor far too large to neutralize the directional bias.

The economic interpretation of the level result is straightforward and arguably the most important contribution of this paper. While much of the existing literature has framed the puzzle as evidence that TIPS are systematically too cheap (typically attributing this to TIPS illiquidity), our results suggest a complementary framing: nominal Treasuries are systematically too expensive, and the reason is that they confer a convenience that TIPS and inflation swaps cannot replicate. A synthetic Treasury constructed from TIPS and inflation swaps replicates the cash flows of a nominal Treasury but not the safety, liquidity, and money-like services that investors derive from holding the nominal bond directly. The wedge between the two should reflect this difference in non-pecuniary value, and the size of the wedge in our sample is consistent with such a story.

The second exercise, in Section VI.B, asks the stronger question of whether variation in the convenience yield explains variation in the mispricing. Here the evidence is meaningfully weaker. In the full sample, the convenience yield explains 18 to 35 percent of the variation in the mispricing, but the slope coefficients are statistically insignificant at the 1- and 2-year tenors. We further find that the intercept is statistically indistinguishable from zero at the 1- and 2-year tenors in the full sample, which can be interpreted as the average mispricing after controlling for the convenience yield not being significantly different from zero, providing further support to the level result.

When we exclude the GFC, the relationship between the convenience yield and the mispricing weakens substantially. The slope coefficients become small and negative, significantly so at the 1-year tenor, and the intercept becomes significantly positive at 20-31 basis points across all tenors. This pattern is not what the convenience yield channel alone would predict. We see two readings of this finding, neither of which we can fully rule out with our data.

The first reading is that the convenience yield acts as a near-constant level shifter outside of crises. Under this reading, the convenience yield is real and is the right size on average to neutralize the directional bias, but it does not vary enough during normal times to drive day-to-day movements in the mispricing. Other factors, such as TIPS liquidity premia, intermediary frictions, and credit risk drive the high-frequency variation, and the convenience yield contributes mostly through its mean. This reading is consistent with Krishnamurthy and Vissing-Jorgensen's (2012) finding that the convenience yield exhibits sharply different behavior in crisis versus non-crisis episodes.

The second reading is more skeptical. It is that variation in the convenience yield is gen-

uinely uninformative about the mispricing outside of crises, and that the strong full-sample co-movement is essentially a single-event phenomenon driven by the joint behavior of these series during 2007–2009. Under this reading, the convenience yield is not a meaningful driver of the puzzle's time-series properties, and the level match in Section VI.A is closer to a useful accident than to evidence of a structural mechanism. We do not think this reading is the most natural one. The magnitude match is informative, and the structural argument in Section III is independent of the empirics, but this reading is consistent with the data and we cannot reject it.

A third possibility, which we view as more likely than either reading above but cannot directly test, is that box-rate measurement error dominates true convenience yield variation outside of crises. NSS estimation noise, residual put-call parity violations, and the use of mid prices as proxies for transaction prices all introduce noise into the constructed convenience yield series. During the GFC, when true convenience yield variation was large, this noise was small relative to signal. During calm periods, the signal-to-noise ratio likely deteriorates substantially, which would attenuate slope coefficients toward zero and could plausibly flip their sign in the presence of any negatively co-moving omitted variable.

We also note that the full-sample slope coefficient at the 1-year tenor exceeds one, which is informative under our structural specification. Equation (3) implies that if the convenience yield were the only relevant friction, the slope of A^1 on r^{conv} should equal one exactly. A slope above one indicates that r^{conv} co-moves positively with another driver of the mispricing – most plausibly a TIPS liquidity premium, which D'Amico et al. (2018) document as peaking during the GFC, the same period during which we and van Binsbergen et al. (2022) document a peak in convenience yields. This positive co-movement during crises is consistent with both convenience yields and TIPS liquidity premia rising together as flight-to-safety phenomena, and it suggests that the full-sample regression captures a combination of the convenience yield channel and other crisis-sensitive factors that we cannot separately identify.

B. Relation to existing literature

Our results contribute to a body of literature that has progressively decomposed the TIPS-Treasury bond puzzle into constituent frictions. Driessen et al. (2017) attribute between 29 and 63 percent of the mispricing to TIPS liquidity, Dittmar et al. (2026) attribute between 46 and 78 percent to U.S. Treasury credit risk, and Ahn and Ahn (2023) attribute around 70 percent to marking-to-market concerns and intermediation frictions combined. Our finding that the convenience yield resolves a substantial portion of the mispricing, particularly its directional component, adds another factor to this list, and one that has been suggested in the literature (Fleckenstein et al., 2014; Pflueger and Viceira, 2016) but not previously tested directly.

The collective explanatory shares in this literature far exceed 100 percent, which means these factors must materially co-move. They are not competing explanations; they are likely partial views of overlapping economic phenomena. Convenience yields rise during flights to safety, as do TIPS liquidity premia and intermediary funding constraints. Treasury credit risk perceptions move with broader macroeconomic stress. Any single-factor regression, including ours, picks up a combination of these correlated drivers and cannot cleanly identify the unique

contribution of any one. A natural next step for this literature is a joint test of these factors in a unified specification. Our paper contributes the framework and the measurement infrastructure to include the convenience yield in such a specification.

Our paper also contributes to the literature on how convenience yields should be measured. van Binsbergen et al. (2022) and Diamond and van Tassel (2022) propose box rates as the optimal reference rate for measuring convenience yields, on the grounds that they are simultaneously risk-free and free of convenience yield. By showing that convenience yields measured this way can resolve the directional component of an entirely different anomaly – one not previously studied by this literature – we provide out-of-sample support for box rates as a useful reference rate. We note, however, that our results also expose a meaningful limitation of the method: the noise in box-rate-based convenience yields is substantial enough that meaningful variation outside of crisis episodes is difficult to detect. Whether this reflects a genuine property of convenience yields (low variation in normal times) or a measurement limitation of box rates is not something we can resolve here, but the question is itself a contribution to the literature on optimal reference rates.

C. Theoretical Implications

Our results suggest that the standard framing of the TIPS-Treasury puzzle as evidence that TIPS are too cheap is incomplete. Reframing the puzzle as evidence that nominal Treasuries are too expensive shifts the locus of the anomaly from the inflation-protected leg of the spread to the nominal leg, and points naturally to convenience yields as a candidate explanation. More broadly, the result implies that empirical work using nominal Treasury yields as a measure of the nominal risk-free rate may carry a systematic downward bias of the order of 25–30 basis points at short tenors. For applications where this bias matters, such as breakeven inflation decompositions, term premium estimation, or any work that subtracts the nominal risk-free rate from another rate, using box rates rather than Treasury yields as the reference rate is likely to produce more economically interpretable results.

D. Practical Implications

For practitioners attempting to harvest the TIPS-Treasury arbitrage, our results imply that the apparent arbitrage is smaller than it appears once the convenience yield foregone by holding the synthetic position is accounted for. The remaining mispricing after this adjustment, while still on average non-zero, is substantially smaller, and is comparable to or smaller than typical bid-ask spreads in TIPS and inflation swap markets. This may help explain why the apparent arbitrage has persisted: a meaningful share of it is not arbitrage at all, but compensation for the convenience that the synthetic position fails to provide.

That being said, accounting for the convenience yield does not eliminate the puzzle. Substantial mispricing remains after accounting for the convenience yield, implying that there are other factors that explain the mispricing. Our work does not contradict previous research in the field. We argue that the convenience yield is one of a multitude of factors explaining the mispricing.

E. Limitations and the Boundary of our Claims

Several limitations bound the interpretation of our results. First, our sample is limited to 1-, 2-, and 3-year tenors, with noted limitations to the extrapolation required for the 1-year TIPS yield. The 1-year results consistently show the largest standard errors and the most volatile coefficient estimates across specifications, which aligns with the literature's view that 1-year TIPS yields are too sensitive to indexation-lags to be wholly interpretable. The 2- and 3-year results are the most reliable, and we view them as carrying most of the weight of our conclusions.

Second, our convenience yield measure is constructed exclusively from SPX options. As shown in Section VII.A, box rates derived from RUT options differ from SPX-derived box rates by approximately 30 basis points on average, a magnitude comparable to the average convenience yield itself. We follow the precedent of van Binsbergen et al. (2022) in using SPX options because of their dominant liquidity, but we cannot rule out that a different choice of underlying would meaningfully alter our results.

Third, we cannot identify the convenience yield channel cleanly from other crisis-sensitive factors. The full-sample regression results in Section VI.B are consistent with convenience yields driving the mispricing during crises, but they are equally consistent with the convenience yield acting as a proxy for a broader stress factor that includes TIPS liquidity premia, intermediary frictions, or credit risk. A joint specification including these factors is needed to make sharper claims, and we view this as the most natural extension of our work.

Fourth, accounting for the convenience yield does not eliminate the puzzle. A statistically significant absolute deviation from no-arbitrage conditions remains in all specifications, and the unexplained portion of the mispricing remains substantial. Our results should be read as identifying the convenience yield as one factor among several, with a particular role in explaining the puzzle's directional component, rather than as a complete resolution of the anomaly.

Taken together, our findings support a calibrated claim: the convenience yield in nominal Treasury securities is real, is approximately the right size to account for why nominal Treasuries are systematically expensive relative to TIPS-plus-inflation-swaps, and belongs in any complete account of the TIPS-Treasury puzzle. Whether convenience yields also drive high-frequency variation in the puzzle outside of crisis episodes is a question on which our evidence is weaker, and which we leave to future work.

IX. Conclusion

This paper investigates whether the convenience yield of nominal Treasury securities can resolve the TIPS-Treasury bond puzzle documented by Fleckenstein et al. (2014). Following van Binsbergen et al. (2022), we measure the convenience yield as the spread between box rates implied by SPX options and nominal Treasury yields, and we test whether incorporating it into the no-arbitrage relationship between TIPS, nominal Treasuries, and inflation swaps reduces the observed mispricing.

We find that adjusting for the convenience yield reduces the average absolute mispricing by 5 to 10 basis points and eliminates the directional bias previously documented in the literature.

The average convenience yield is of approximately the right size to account for why nominal Treasuries are systematically expensive relative to TIPS-plus-inflation-swaps. We interpret this as evidence that the puzzle is, in part, not a puzzle: a meaningful share of the wedge reflects compensation for the convenience that a synthetic Treasury constructed from TIPS and inflation swaps cannot replicate.

The convenience yield does not, however, explain time-series variation in the mispricing outside of crisis episodes. Our regression results are dominated by the Great Financial Crisis, and excluding it leaves the relationship between the convenience yield and the mispricing weak and wrong-signed. We read this as consistent with the convenience yield acting primarily as a level shifter in normal times, with measurement noise in box rates plausibly obscuring whatever variation exists.

Our contribution is therefore narrower than a full resolution of the puzzle but specific in what it adds: convenience yields belong in the conversation alongside TIPS liquidity premia, Treasury credit risk, and intermediation frictions, and they specifically account for the puzzle's most distinctive feature, its consistent direction. A natural next step is to test these factors jointly, since their individually estimated explanatory shares far exceed 100 percent and they almost certainly co-move. We leave this to future work.

References

Ahn, Jungkyu, and Yongkil Ahn, 2023, What drives the TIPS–Treasury bond mispricing?, Journal of Empirical Finance 74.

Andreasen, Martin M., Jens H. E. Christensen, and Simon Riddell, 2021, The TIPS liquidity premium, Review of Finance 25, 1639–1676.

Ashton, Michael, 2008, Drivers of inflation-linked corporate bond spreads and the inflation swap/bond breakeven difference, working paper, SSRN.

Augustin, Patrick, Mikhail Chernov, Lukas Schmid, and Dongho Song, 2021, Benchmark interest rates when the government is risky, Journal of Financial Economics 140, 74–100.

Brunnermeier, Markus K., 2009, Deciphering the liquidity and credit crunch 2007–2008, Journal of Economic Perspectives 23, 77–100.

Campbell, John Y., Robert J. Shiller, and Luis M. Viceira, 2009, Understanding inflation-indexed bond markets, Brookings Papers on Economic Activity 2009, 79–120.

Choy, Siu Kai, and Jason Wei, 2020, Liquidity risk and expected option returns, Journal of Banking and Finance 111.

Christoffersen, Peter, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui, 2018, Illiquidity premia in the equity options market, Review of Financial Studies 31, 811–851.

Cleveland, Robert B., William S. Cleveland, Jean E. McRae, and Irma Terpenning, 1990, STL: A seasonal-trend decomposition procedure based on Loess, Journal of Official Statistics 6, 3–73.

D'Amico, Stefania, Don H. Kim, and Min Wei, 2018, TIPS from TIPS: The informational content of Treasury Inflation-Protected Security prices, Journal of Financial and Quantitative Analysis 53, 395–436.

Diamond, William, and Peter van Tassel, 2022, Risk-free rates and convenience yields around the world, IDEAS working paper.

Dittmar, Robert F., Alex Hsu, Guillaume Roussellet, and Peter Simasek, 2026, Default risk and the pricing of U.S. sovereign bonds, Journal of Finance 81, 829–869.

Driessen, Joost, Theo E. Nijman, and Zorka Simon, 2017, The missing piece of the puzzle: Liquidity premiums in inflation-indexed markets, SAFE Working Paper No. 183, Goethe University Frankfurt.

Du, Wenxin, Joanne Im, and Jesse Schreger, 2018, The U.S. Treasury premium, Journal of International Economics 112, 167–181.

Findley, David F., 2005, Some recent developments and directions in seasonal adjustment, Journal of Official Statistics 21, 343–365.

Findley, David F., Brian C. Monsell, William R. Bell, Mark C. Otto, and Bor-Chung Chen, 1998, New capabilities and methods of the X-12-ARIMA seasonal-adjustment program, Journal

of Business and Economic Statistics 16, 127–152.

Fleckenstein, Matthias, Francis A. Longstaff, and Hanno Lustig, 2014, The TIPS-Treasury bond puzzle, Journal of Finance 69, 2151–2197.

Fleming, Michael J., 2003, Measuring Treasury market liquidity, Economic Policy Review (Federal Reserve Bank of New York) 9, 83–108.

Fleming, Michael J., Bruce Mizrach, and Giang Nguyen, 2018, The microstructure of a U.S. Treasury ECN: The BrokerTec platform, Journal of Financial Markets 40, 2–22.

Gregory, Allan W., Bruce E. Hansen, 1996, Residual-based tests for cointegration in models with regime shifts, Journal of Econometrics 70, 99–126.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2007, The U.S. Treasury yield curve: 1961 to the present, Journal of Monetary Economics 54, 2291–2304.

Gürkaynak, Refet S., Brian Sack, and Jonathan H. Wright, 2010, The TIPS yield curve and inflation compensation, American Economic Journal: Macroeconomics 2, 70–92.

Hampel, Frank R., 1974, The influence curve and its role in robust estimation, Journal of the American Statistical Association 69, 383–393.

Haubrich, Joseph, George Pennacchi, and Peter Ritchken, 2012, Inflation expectations, real rates, and risk premia: Evidence from inflation swaps, Review of Financial Studies 25, 1588–1629.

Hu, Grace Xing, Jun Pan, and Jiang Wang, 2013, Noise as information for illiquidity, Journal of Finance 68, 2341–2382.

Kalman, R. E., 2011, A new approach to linear filtering and prediction problems, in Robert Fildes, and P. Geoffrey Allen, eds.: Forecasting (SAGE Publications, London).

Krishnamurthy, Arvind, 2002, The bond/old-bond spread, Journal of Financial Economics 66, 463–506.

Krishnamurthy, Arvind, and Annette Vissing-Jorgensen, 2012, The aggregate demand for Treasury debt, Journal of Political Economy 120, 233–267.

Longstaff, Francis A., 2004, The flight-to-liquidity premium in U.S. Treasury bond prices, Journal of Business 77, 511–526.

Nagel, Stefan, 2016, The liquidity premium of near-money assets, Quarterly Journal of Economics 131, 1927–1972.

National Bureau of Economic Research, 2010, Business cycle dating committee announcement, September 20, 2010, Cambridge, MA, https://www.nber.org/news/business-cycle-dating-committee-announcement-september-20-2010.

Nelson, Charles R., Andrew F. Siegel, 1987, Parsimonious Modeling of Yield Curves, The Journal of Business 60, 473–489.

Perron, Pierre, 1989, The great crash, the oil price shock, and the unit root hypothesis, Econometrica 57, 1361–1401.

Pflueger, Carolin E., and Luis M. Viceira, 2016, Return predictability in the Treasury market: Real rates, inflation, and liquidity, in Pietro Veronesi, ed.: Handbook of Fixed-Income Securities (John Wiley and Sons, Hoboken, NJ), 191–209.

Schwartz, Henry, 2026, The state of the options industry: 2025, Cboe Global Markets, January 22, https://www.cboe.com/insights/posts/the-state-of-the-options-industry-2025/.

Stock, James H., 1987, Asymptotic properties of least squares estimators of cointegrating vectors, Econometrica 55, 1035–1056.

Svensson, Lars E.O., 1994, Estimating and interpreting forwards interest rates: Sweden 1992 - 1994, NBER Working Paper No. 4871.

van Binsbergen, Jules H., William F. Diamond, and Marco Grotteria, 2022, Risk-free interest rates, Journal of Financial Economics 143, 1–29.

Wright, Jonathan H., 2018, Seasonal adjustment of NIPA Data, NBER Working paper 24895.

Appendix

Appendix A: The Process of NSS-Estimation of Yield Curves

NSS-estimation is used in two places: to fit a daily box-rate yield curve from the panel of tenor-specific box rates obtained from SPX options, and to observe synthetic 1-, 2-, and 3-year TIPS yields from the daily NSS parameters published by the Federal Reserve. Both applications rely on the same parametric form, which we review briefly here. The model by Nelson and Siegel (1987) specifies the zero-coupon yield at tenor n as a four-parameter function:

$$y(n) = \beta_0 + \beta_1 \left(\frac{1 - e^{-n/\tau_1}}{n/\tau_1} \right) + \beta_2 \left(\frac{1 - e^{-n/\tau_1}}{n/\tau_1} - e^{-n/\tau_1} \right)$$

The three terms admit a standard economic interpretation: β_0 governs the long-run level, β_1 the short-end slope (its loading decays monotonically from one to zero as n grows), and β_2 a medium-term hump whose location is controlled by the decay parameter $\tau_1 > 0$. Svensson (1994) extends the model by adding a second curvature term with its own decay parameter τ_2:

$$y(n) = \beta_0 + \beta_1 \left(\frac{1 - e^{-n/\tau_1}}{n/\tau_1} \right) + \beta_2 \left(\frac{1 - e^{-n/\tau_1}}{n/\tau_1} - e^{-n/\tau_1} \right) + \beta_3 \left(\frac{1 - e^{-n/\tau_2}}{n/\tau_2} - e^{-n/\tau_2} \right)$$

The additional term allows the fitted curve to accommodate a second hump or trough, which matters at longer horizons where a single-hump Nelson–Siegel form tends to misfit observed Treasury curves (Gürkaynak, Sack, and Wright, 2007). We refer to the six-parameter vector as the NSS parameters.

$$\theta = (\beta_0, \beta_1, \beta_2, \beta_3, \tau_1, \tau_2)$$

For the box-rate curve, we estimate θ separately on each trading day by non-linear least squares:

$$\theta_t = \arg\min_{\theta} \sum_i \left[r^{box}_{t,n_i} - y(n_i; \theta) \right]^2$$

where the sum runs over all expiry groups i for which a box rate is observable on day t, and $y(n_i; \theta)$ is the NSS yield at tenor n_i evaluated at θ. Because the objective is non-convex in (τ_1, τ_2), we use multi-start optimization with starting values drawn from a grid over the decay parameters and retain the best-fitting local optimum. Once θ_t is obtained, the fitted box rate at any tenor n is simply:

$$r^{box}_{t,n} = y(n, \theta_t)$$

For TIPS yields we do not perform the fit ourselves. The Federal Reserve publishes daily NSS parameters estimated from the cross-section of outstanding TIPS, and we use these published parameters directly to compute $r^{TIPS}_{t,n}$ at our target tenors.

Two caveats apply. First, NSS is a smoothing device, not a no-arbitrage term-structure model; fitted yields at any single tenor inherit information from the full cross-section of observed rates on that day, which is a feature when inputs are noisy but a limitation when the fitted tenor lies

far outside the range of observed maturities. Second, because $n_i = 0$ makes the Nelson–Siegel loadings indeterminate, we evaluate $\lim_{n\to 0}(1 - e^{-n/\tau})/(n/\tau) = 1$ when needed.

Appendix B: Inspection of Inflation-Seasonality

In theory, inflation swaps and TIPS yields of the same tenor should react to inflation-seasonality in perfect inversion to one another. As can be seen in Figure B1, however, the magnitude of seasonality in the TIPS yield is much larger than the magnitude of seasonality in the inflation swap. As explained in Section V.A, this is caused by a mismatch in the actual tenors of the contracts with which the 1-, 2-, and 3-year tenored interest rates are estimated for the TIPS yield and inflation swap rate respectively.

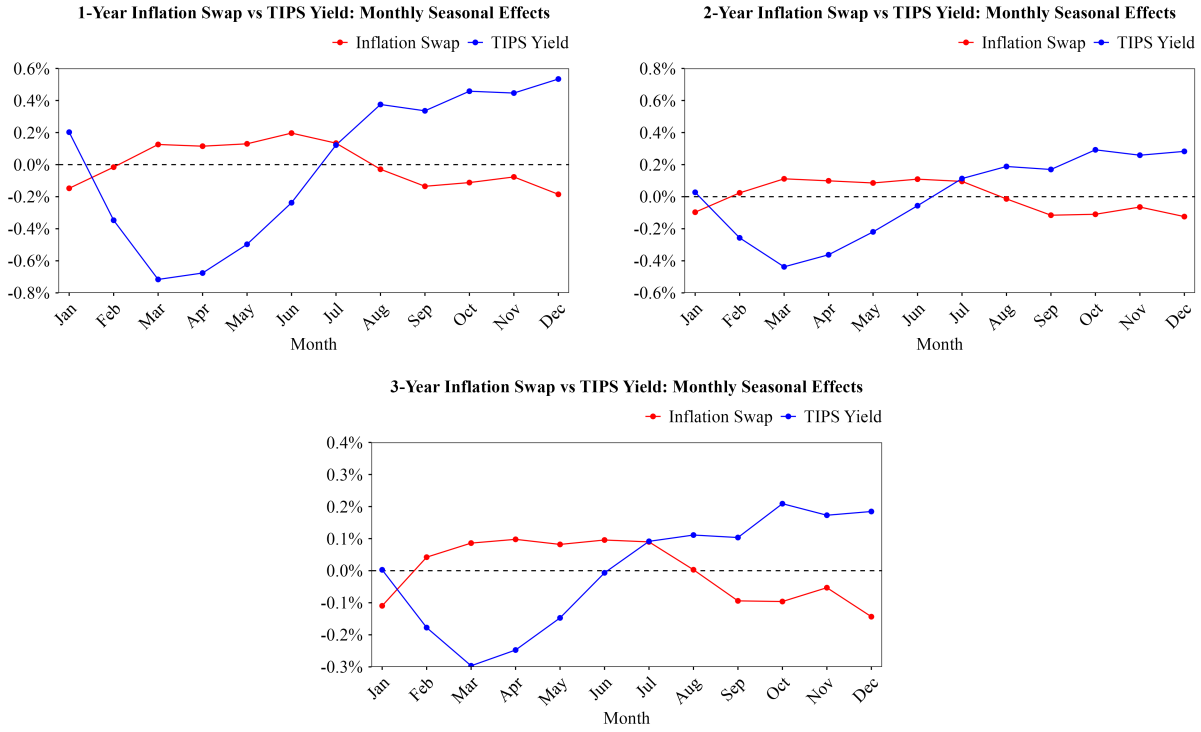


Figure B1. *Monthly Seasonal Effects of Inflation Swaps and TIPS Yields, at 1-Year, 2-Year, and 3-Year Tenors, Full Sample*

This causes seasonality in A^1 as shown below. We perform a HAC F-test for monthly seasonality with the null hypothesis that no monthly seasonality exists. The null hypothesis is strongly rejected across all three tenors of A^1.

Table B1. *HAC F-Test for Monthly Seasonality in the Mispricing A^1, at 1-Year, 2-Year, and 3-Year Tenors, Full Sample*

Statistic	1-Year	2-Year	3-Year
F-statistic	31.0131	12.7564	2.8494
p-value	0.0000	0.0000	0.0010

To confirm that applying seasonality-adjustments to A^1 and A^2 serves only to resolve this

inflation-seasonality mismatch and does not strip our data of any economically valuable seasonality in the nominal Treasury security, we inspect the seasonality of the nominal Treasury security. As can be seen below, the nominal Treasury security does not exhibit any seasonality.



Figure B2. *Monthly Seasonal Effects of Nominal Treasury Yield at 1-Year, 2-Year, and 3-Year Tenors, Full Sample*

Appendix C: AI-Disclosure

In the writing of this thesis, we have used AI for three purposes: searching for literature, refining our writing, and assisting in programming. The models used are Claude and ChatGPT. For literature search, AI was used to efficiently find relevant papers on specified topics. We verified all suggested citations to confirm the correct attribution and cited content. For writing, AI was used to receive suggestions on how to clarify ideas and improve readability. All intellectual arguments and conclusions are our own. Lastly, for programming, AI was used to assist in writing and debugging code. All suggestions made by AI have been tested and reviewed to ensure theoretical accuracy and correct functionality.