Discussion Paper #0024

September 2023

Tests for Jumps in Yield Spreads

Lars Winkelmann
Wenyong Yao
Tests for Jumps in Yield Spreads

Lars Winkelmann∗
Department of Economics, Freie Universität Berlin
and
Wenying Yao†
Melbourne Business School, University of Melbourne

September 1, 2023

Abstract

This paper studies high-frequency econometric methods to test for a jump in the spread of bond yields. We propose a coherent inference procedure that detects a jump in the yield spread only if at least one of the two underlying bonds displays a jump. Ignoring this inherent connection by basing inference only on a univariate jump test applied to the spread tends to overestimate the number of jumps in yield spreads and puts the coherence of test results at risk. We formalize the statistical approach in the context of an intersection union test in multiple testing. We document the relevance of coherent tests and their practicability via simulations and real data examples.

Keywords: High-frequency data; jumps; sequential testing; intersection union test; term spread; break-even inflation.

JEL: C58, C12, E43, E44

∗The authors are grateful to the Editor, the Associate Editor, and two anonymous referees for helpful and constructive comments. Winkelmann acknowledges financial support from the Deutsche Forschungsgemeinschaft (DFG) project number 363880538. Winkelmann and Yao also acknowledge financial support from the Australian Research Council Discovery Grants Program (DP220100321).

†Corresponding author: Melbourne Business School, University of Melbourne, 200 Leicester Street, Carlton VIC 3053, Australia; Phone: +61 (3) 9349 8141; Email: W.Yao@mbs.edu.
1 Introduction

News announcements generate significant discontinuities (jumps) in financial asset prices and reveal important information about expectations and risks. The prevailing presence and high information content of jumps have been demonstrated by the rich literature on jumps in stock prices (Lee and Mykland, 2008; Lee, 2012), bond yields (Jiang et al., 2011; Winkelmann et al., 2016) and currencies (Scaillet et al., 2020; Lee and Wang, 2020). However, the interplay in the jump behaviour of a pair of assets and the spread between them at high-frequency is less known. In particular, yield spreads between a pair of bonds contain economically relevant information. Negative term spreads indicate recessions (Henry and Phillips, 2020; Yang, 2020), rising credit spreads inform about increasing default risks (Del Negro and Schorfheide, 2013; Leombroni et al., 2021), and the yield spread between nominal and inflation-indexed government bonds, commonly known as break-even inflation, is a market-based measure of inflation expectation (Chernov and Mueller, 2012; Hanson and Stein, 2015).

A natural approach to test for jumps in yield spreads is pursued by Boffelli and Urga (2015). The authors consider credit spreads as an observable variable and apply a jump test directly on the spread. However, investigating jumps in spreads between financial assets is not as straightforward as it might seem, as an inherent multiplicity complicates the statistical inference. Within the general class of continuous-time semimartingales with Brownian and jump components, the yield spread jumps only if at least one of the underlying bond yield processes jump contemporaneously. In practice, such coherent test results are not guaranteed. Univariate jump tests applied to the yield spread may detect a jump while it is possible that the same tests detect no jump when applied to the two individual bond yields. As an example, Figure 1 shows confidence sets of the Lee and Mykland (2012) jump test applied to the 10-year break-even inflation rate (diagonal corridor) and the underlying
Figure 1: Confidence sets of Lee-Mykland jump tests on two bond yields and the yield spread at Initial Jobless Claims release time on Oct 24, 2019.

Note: We implement the Lee-Mykland test (Lemma 1 of Lee and Mykland, 2012) on 30-second bond yields data at 8:30 a.m. EST. ‘×’ shows the estimated bond yield changes. The diagonal corridor outlines the 99% confidence set for no jump in the yield spread. The square corresponds to the confidence set of Bonferroni-adjusted Lee-Mykland tests applied to the two bond yields with a family-wise error rate of 1%. The dotted areas are where incoherent test outcomes occur, i.e., a jump in the yield spread is detected but no jump in the two bond yields. See Section 4.1 for more details of the underlying data.

nominal and inflation-indexed bond yields (square) on October 24, 2019, when the U.S. Department of Labor Statistics released data on Initial Jobless Claims. The two dotted areas correspond to pairs of contemporaneous yield changes with conflicting test results. That is, neither of the two bond yields are tested to contain a jump, but the spread between them displays evidence of a jump. The estimated yield change at the news release time is marked as a red cross in Figure 1, which falls into the area of incoherent test results.

The incoherency is a direct consequence of the fact that multiple nested hypotheses are involved when testing for jumps in yield spreads. It is therefore not a feature exclusive to the Lee-Mykland test, but a property shared by all univariate jump tests applied to spreads. Furthermore, conflicting test results are not limited to finite-sample but persist asymptotically. Coherence is one of the most important concepts in the multiple testing literature (Finner and Strassburger, 2002), which motivates the so-called closed or stepwise test procedures. The closure principle of Marcus et al. (1976) is a well established criterion in different areas of application, including clinical trails and multiple endpoint studies, see
Liu and Hsu (2009) for example.

This paper proposes coherent tests for jumps in yield spreads. We study intersection union tests (IUT) that are based on two elementary hypotheses: (i) the hypothesis of no jump in both bond yields; and (ii) the hypothesis of equal yield changes in the two bonds. The first step uses a bivariate jump test to test for a jump in at least one of the two underlying bond yields. If the first step is rejected, we proceed with the second step to test the equality of the two yield changes. The rejection of both elementary hypotheses detects a jump in the spread. The proposed IUT is a special case of step-wise tests that are nested in sequence, and hence controls the joint error rate without multiplicity adjustment, see Hsu and Berger (1999) and further generalizations by Goeman and Solari (2010). The IUT procedure fulfills the closure principle, which guarantees coherent test results.

Univariate tests for the presence of jumps in discretely observed semimartingale models have been proposed by Aït-Sahalia and Jacod (2009) and Lee and Mykland (2008), for example. Dumitru and Urga (2012) provide an overview and empirical comparison of existing tests. Hansen and Lunde (2006) highlight that trading frictions, such as price discreteness and bid-ask bounces, play a non-trivial role at high observation frequencies. Commonly referred to as the market microstructure noise, these frictions keep observed prices and yields away from discretely observed semimartingales, and distort simple estimators of volatility and jumps. Therefore, Podolskij and Ziggel (2010), Aït-Sahalia et al. (2012), Lee and Mykland (2012) and Bibinger et al. (2019) provide noise-robust, univariate jump tests in latent observation models. Extensions of the univariate methods to contemporaneous jumps (cojumps) in bivariate semimartingale models are proposed by Jacod and Todorov (2009) and Bibinger and Winkelmann (2015).

The first step of the IUT can be accomplished in two different ways. The first approach, used in the construction of Figure 1, implements univariate jump tests on each of the two bond yields with a Bonferroni correction. The Bonferroni approach is conservative in
detecting a jump, and can be improved especially when the joint distribution of changes in bond yields under the null is available. We propose a \( \chi^2 \)-jump test as a more powerful alternative by extending the local pre-averaging jump test of Lee and Mykland (2012) to the bivariate case. While Lee and Mykland (2012) use a jump-diffusion model, their local jump test has been generalized by Bibinger et al. (2019) to semimartingale models with stochastic volatility and leverage. We use the same setup as Bibinger et al. (2019) but in the two-dimensional space. A feasible central limit theorem for changes in a pair of bond yields at a given event time under the hypothesis of no jumps forms the basis of the \( \chi^2 \)-jump test. Bivariate jump tests are conceptually different to cojump tests, because they have power against jumps in only one of the two bonds.

The limiting distribution of the two bond yields is also used in constructing the test for equal yield changes in the second step of the IUT. Since the spread is the difference of the two underlying bond yields, testing for equal yield changes is the same as testing the spread for a jump. Therefore, the second step of the IUT applies the Lee-Mykland jump test to the spread. The two-step IUT procedure makes use of the information on the two underlying bond yields when testing for jumps in the spread. In contrast, a single-step univariate jump test on the spread ignores information on the two bond yields and treats all combinations of bond yield changes that result in the same yield spread equally.

We quantify the benefit of guarding against incoherency via simulations. Across different significance levels and correlations between the two bond yields, the Bonferroni-based IUT reveals that up to 90% of the falsely detected spread jumps by the univariate procedure are incoherent. The percentage reduces to around 50% for the \( \chi^2 \)-based approach. Although asymptotically, incoherent test results can only occur under the global null hypothesis, simulations show that in finite sample, incoherent results also occur under the alternative of a jump in the yield spread with non-negligible probabilities. While guarding against incoherency appears desirable under the global null hypothesis of no jumps, it is
costly under the alternative as it reduces the power of the jump test on the spread. By the consistency of the elementary tests, such costs in the form of power losses vanish in large sample while the benefit of coherent test results remains.

We demonstrate the spread jump tests using high-frequency data on U.S. government bond yields from 2017 to 2019. High-frequency data of U.S. government bonds have been studied by Jiang et al. (2011), Hördahl et al. (2020), and the references therein. In contrast to previous papers, we study the occurrence of conflicting test results by focusing on term spreads and break-even inflation rates at major macroeconomic news release times. Using the same Bonferroni-based IUT procedure as in Figure 1, up to 25% of the locally detected spread jumps do not have corresponding jumps in the two underlying bond yields. Such probabilities decrease to around 5% when using the $\chi^2$-based IUT. The empirical results confirm that the proposed IUT based on the $\chi^2$-jump test appears to be a good compromise in establishing coherence of the test at a minimum reduction in the detection of spread jumps.

The remainder of the paper is structured as follows. Section 2 introduces the econometric model for the bond yields and the tests for a jump in the yield spread. Section 3 compares the IUT and the Lee-Mykland test applied to the yield spread via simulations. Section 4 demonstrates the usefulness of the proposed IUT using high-frequency bond yields data. Finally, Section 5 concludes. Detailed technical assumptions and proofs are delegated to the online appendix, as well as additional simulation results and empirical analyses.

2 Econometric method

This section introduces the theoretical framework to investigate jumps in two bond yields and their spread at some fixed point in time using intra-day data observed with market microstructure noise.
2.1 Underlying stochastic process

Let $y_t$ denote the time-$t$ vector of bond yields that relates to the efficient prices of some bond $a$ and bond $b$.\(^1\) We consider $y_t$ on a normalized interval with some fixed start and end time, surrounding a fixed and exogenous event time $\tau$. In our empirical application, $\tau$ is determined by some pre-scheduled macroeconomic news release. The bond yields can be described by a continuous-time bivariate Itô semimartingale:

$$y_t = y_0 + \int_0^t b_s ds + \int_0^t \sigma_s dW_s + J_t, \quad t \in [0, 1]. \quad (1)$$

The continuous part consists of the starting value $y_0$, a two-dimensional drift $b_t$, the $2 \times 2$ spot-covolatility $\Sigma_t = \sigma_t \sigma_t'$, and a two-dimensional standard Brownian motion $W_t$. $J_t$ is a purely discontinuous process that can be completely characterized by its jumps. Assumptions on the different components of $y_t$ are further formalized in Appendix (A). Model (1) is fairly general and includes most models for asset prices in financial econometrics, particularly those introduced by Duffie and Kan (1996) for bond yield processes. The occurrence of jumps is not restricted to the event time $\tau$, but can be distributed anywhere on $[0, 1]$. The statistical methods below remain valid, provided that the estimation of covolatility $\Sigma_t$ is robust to such jumps at $t \neq \tau$.

If we were able to observe $y_t$ in continuous time, we could have observed all jumps directly. The return $\Delta y_t = y_t - y_{t-}$, $y_{t-} = \lim_{s < t, s \to t} y_s$ is zero in the case of no jump at time $t$, and $\Delta y_t = \Delta J_t$ otherwise. However, in any practical applications we have only finitely many observations of the two bond yields. We consider observation times $t_j$, $j = 1, ..., n$, that are discrete, synchronous, and equally spaced across bonds. The sampling interval $t_j - t_{j-1}$ has length $1/n$. Besides the finite $n$, we follow most of the market microstructure

\(^1\)For a zero coupon bond the relationship between yields, $y_t$, and prices, $P_t$, is $y_t^{(i)} = -\log P_t^{(i)}/m^{(i)}$, with time to maturity $m^{(i)}$, $i = a, b$. Detecting jumps in the bond yields and bond prices are interchangeable. Given the focus of the paper, we refer to yields directly instead of log-prices.
literature and posit an additive, latent observation model in discrete time:

$$\tilde{y}_j = y_{t_j} + \epsilon_j.$$  \hspace{1cm} (2)

$\tilde{y}_j$ is the observed noisy version of the efficient process $y_{t_j}$, where $\epsilon_j \sim (0, \eta)$, $j = 1, \ldots, n$, is the market microstructure noise with a $2 \times 2$-dimensional covariance matrix $\eta$. Aït-Sahalia and Yu (2009) find a negative relationship between the level of the noise variance and different liquidity measures. Less liquid assets usually have larger noise variance. Distorting effects through potential differences in the liquidity of bonds are captured by the noisy observation model (2). Methods to test for significant noise are proposed by Aït-Sahalia and Xiu (2019), for example. Similar to the jump component, increments of the noise term do not vanish asymptotically. Hence, covolatility estimators that are not robust to market microstructure are asymptotically dominated by the noise. Jump detection is more complicated in cases where jump returns are weakened by the noise while no-jump returns are amplified. Lee and Mykland (2012) highlight the importance of noise-robust jump tests compared to methods that do not account for market microstructure noise. The market microstructure is assumed to be an \textit{i.i.d.} process independent of $y_t$. We extend to a more general setup in the simulation with endogenous and heteroskedastic noise.

### 2.2 Bivariate distribution of pre-averaged event returns

Given the noisy observation model (2), smoothing the observed yields is the natural approach to diminish the impact of market microstructure noise. Following the general pre-averaging approaches of Podolskij and Vetter (2009), Jacod et al. (2009) and Christensen et al. (2010), we use the average of the observed yields over $M_n$ discrete noisy observations $\tilde{y}_j = (\tilde{y}_j^{(a)}, \tilde{y}_j^{(b)})$,

$$\hat{y}_j = M_n^{-1} \sum_{i=j}^{(j+M_n-1)/n} \tilde{y}_i, \quad j = 1, \ldots, n,$$  \hspace{1cm} (3)
where the block length $M_n = c\sqrt{n}$ is an integer, and $c$ is a constant tuning parameter. The block length balances the orders of the noise and continuous component of the efficient yields. Since the microstructure noise is centered and serially uncorrelated, taking averages of the noisy observations reduces the impact of the noise component. As a result, the estimated return vector at announcement time $\tau$,

$$\Delta \hat{y}_{[\tau n]} = \hat{y}_{[\tau n]} - \hat{y}_{[\tau n]-M_n},$$ (4)

is close to the latent increment of the yield $\Delta y_\tau$, and is no longer dominated by the noise. The index $[\tau n]$ denotes the smallest integer larger than $\tau n$, and accounts for the case where the announcement time does not exactly match with the observation times of the processes.

The following proposition describes the limiting distribution of the estimated event return of the two bonds $\Delta \hat{y}_{[\tau n]} = (\Delta \hat{y}_{[\tau n]}^{(a)}, \Delta \hat{y}_{[\tau n]}^{(b)})'$. It is a direct extension of Proposition 3.1 of Bibinger et al. (2019) from the univariate to the bivariate case.

**Proposition 2.1** In the model presented in Section 2.1 under Assumptions 1 and 2 in Appendix (A), the return vector (4) of the pre-averaged yields satisfies

$$n^{1/4} (\Delta \hat{y}_{[\tau n]} - \Delta y_\tau) \xrightarrow{(st)} MN(0, \Gamma),$$ (5)

where $MN$ denotes a mixed normal distribution, and the covariance matrix $\Gamma$ has elements

$$\Gamma^{(i,j)} = \frac{1}{3} \left( \Sigma^{(i,i)}_{\tau} + \Sigma^{(i,j)}_{\tau} - \Sigma^{(j,j)}_{\tau} \right) c + 2c^{-1}\eta^{(i,j)}, \quad i, j = a, b.$$ (6)

Proposition 2.1 shows that the simple pre-averaging approach consistently estimates the event return. The limiting spot variances $\Sigma^{(i,i)}_{\tau}$, $i = a, b$ and covariance $\Sigma^{(a,b)}_{\tau}$ can be random (in the case of stochastic volatility, for example), and therefore the limiting distribution in (5) is a mixed normal. The return estimator has the optimal rate of $n^{1/4}$, which is a direct consequence of the choice of the block length $M_n$ and the order of the continuous components of the yield $y_t$. As the variance-covariance matrix in (6) indicates, the diagonal and off-diagonal elements of $\Gamma$ have the same structure, provided that the microstructure
noise displays a non-zero covariation between bonds $a$ and $b$, $\eta^{(a,b)} \neq 0$. $\Gamma$ accounts for contemporaneous jumps in (co)volatility by referring to spot (co)volatility before ($\Sigma_{\tau-}$) and at ($\Sigma_{\tau}$) the event time.\footnote{See, for example, Bibinger and Winkelmann (2018) and Li et al. (2021), for studies on jumps in (co)volatility.}

Stable convergence in law ($st$) and a consistent estimator of the covariance $\hat{\Gamma}$ provide a feasible, self-normalizing version of (5), which we will exploit below. Compatible estimators for integrated (co)volatility and noise variance in the present modelling context are proposed by Christensen et al. (2010) and Koike (2016).

### 2.3 Tests for jumps in bond yields

To study whether at least one of the two bond yields $y_t = (y_t^{(a)}, y_t^{(b)})'$ jumps at event time $t = \tau$, we aim at distinguishing the (point) hypothesis $H_B^0(\tau) : \Delta y_\tau = 0$ from the alternative $H_B^1(\tau) : \Delta y_\tau \neq 0$, where the superscript $B$ refers to the two bond yields. Under $H_B^1(\tau)$, the yields of either bond $a$ or bond $b$, or both jump at time $\tau$. We consider two ways to approach this testing problem.

The first approach is to use the local Lee-Mykland jump test in Proposition 3.1 of Bibinger et al. (2019) on the two bond yields separately with a Bonferroni adjustment:

$$
\varphi_B^\alpha(\text{Bonf}) = \{ n^{1/4} \frac{\Delta y^{(a)}_{\lceil \tau n \rceil}}{\sqrt{\hat{\Sigma}^{(a,a)}_{\lceil \tau n \rceil}}} > q_{1-\alpha/4}(N) \cup n^{1/4} \frac{\Delta y^{(b)}_{\lceil \tau n \rceil}}{\sqrt{\hat{\Sigma}^{(b,b)}_{\lceil \tau n \rceil}}} > q_{1-\alpha/4}(N) \}.
$$

We decide in favor of $H_B^1(\tau)$, if at least one of the univariate test statistics is larger than the $(1 - \alpha/4)$-quantile of the standard normal distribution, $q_{1-\alpha/4}(N)$. While the Bonferroni approach controls the family-wise error rate at $\alpha$, it ignores the correlation between the two test statistics. Thus, it is conservative in testing $H_B^0(\tau)$, especially when the event returns of the two bonds are highly correlated.

The second approach improves upon the Bonferroni test using the bivariate distribution of event returns from Proposition 2.1.
Corollary 2.2 Given Proposition 2.1, under $H^B_0(\tau)$, the bivariate $\chi^2$-jump test at the event time $\tau$,

$$\varphi^B_\alpha(\chi^2) = 1 \left\{ \sqrt{n} \Delta y'_{[\tau_n]} \hat{\Gamma}^{-1}_\tau \Delta y_{[\tau_n]} > q_{1-\alpha}(\chi^2(2)) \right\}, \quad (8)$$

is an asymptotic level-$\alpha$ test. $q_{1-\alpha}(\chi^2(2))$ denotes the $(1-\alpha)$-quantile of the $\chi^2$ distribution with two degrees of freedom. The test is consistent under the alternative $H^B_1(\tau)$ with divergence rate $n^{1/2}$.

Corollary 2.2 shows a standard result in multivariate statistics for the sum of squares of whitened event returns, $z_\tau = \hat{\Gamma}^{-1/2}_\tau \Delta y_{[\tau_n]}$. Under $H^B_0(\tau)$, elements of the $2 \times 1$ vector $n^{1/4}z_\tau$ are independent standard normal random variables. As a result, the sum of squares $\sqrt{n} z'_\tau z_\tau$ in (8) has an asymptotic $\chi^2$ distribution with two degrees of freedom. The $\sqrt{n}$ convergence rate is a direct consequence of the $n^{1/4}$ rate of pre-averaged returns in Proposition 2.1. The $\chi^2$-jump test is efficient in the sense that its $(1-\alpha)$-confidence set spans the smallest area in the two-dimensional event-return space among all level-$\alpha$ tests of $H^B_0(\tau)$.

2.4 Tests for jumps in yield spreads

The test for a jump in the yield spread can be translated into the composite hypothesis

$$H^S_0(\tau) : \Delta y^{(a)}_{\tau} - \Delta y^{(b)}_{\tau} = 0 \quad \text{against} \quad H^S_1(\tau) : \Delta y^{(a)}_{\tau} - \Delta y^{(b)}_{\tau} \neq 0,$$

where the superscript $S$ refers to the spread. Under the null hypothesis, yield changes of bond $a$ and bond $b$ at time $\tau$ are identical, which implies that either both bonds do not jump or cojump with the same jump size. Therefore, $H^S_0(\tau)$ is equivalent to no jump in the yield spread at time $\tau$. The alternative hypothesis states that there is a jump in the spread at time $\tau$. It reveals a necessary condition for a jump in the yield spread—a jump in the spread occurs only if at least one of the underlying bond yields has a jump: $\Delta y^{(i)}_{\tau} \neq 0$ for $i = a$ or $i = b$. This relationship imposes a multiple testing problem, since evidence supporting $H^S_1(\tau)$ should be confirmed by rejecting $H^B_0(\tau)$.
The first and probably most natural approach to test $H_0^S(\tau)$ in our model context is the Lee-Mykland test applied to the spread. Using Proposition 2.1, we obtain the test function

$$\varphi^S_\alpha(LM) = 1 \left\{ n^{1/4} \left| \frac{\Delta \hat{y}_{[\tau n]}^{(a)} - \Delta \hat{y}_{[\tau n]}^{(b)}}{\sqrt{\hat{\Gamma}^{(a,a)}_\tau + \hat{\Gamma}^{(b,b)}_\tau - 2 \hat{\Gamma}^{(a,b)}_\tau}} \right| > q_{1-\alpha/2}(N) \right\}. \quad (9)$$

The test statistic has the change in the spread at the event time in its numerator and the standard deviation of the change in the spread in the denominator. Notice that (9) is analogous to the difference-in-differences approach, where we test whether some “treatment” at time $\tau$ results in significantly different responses of the two bonds. While $\varphi^S_\alpha(LM)$ is a consistent level-$\alpha$ test, the following corollary states that it is not coherent.

**Corollary 2.3** Assume Proposition 2.1 and the validity of the global null hypothesis $H_{0B}^R(\tau)$. Then, for significance levels $0 < \alpha < 1$, the test $\varphi^S_\alpha(LM)$ is not coherent for $H_0^S(\tau)$. That is, for any level-$\alpha$ test $\varphi^B_\alpha$ of $H_{0B}^R(\tau)$, the impossible event of $H_{0B}^R(\tau) \cap H_1^S(\tau)$ is detected with non-zero probability: $P(\varphi^B_\alpha = 0, \varphi^S_\alpha(LM) = 1) > 0$.

The non-zero probability in Corollary 2.3 is an immediate consequence from testing a specific point in the event-return space, $H_{0B}^R(\tau)$, jointly with a set of values, $H_0^S(\tau)$, at the same level $\alpha$, while the former is a special case of the latter. The probability of conflicting test results under the hypothesis of no jumps in Corollary 2.3 can be expressed as integrals of bivariate normal distributions when replacing $\varphi^B_\alpha$ by either $\varphi^B_\alpha(\chi^2)$ or $\varphi^B_\alpha(Bonf)$ from Section 2.3. Appendix (A) provides the derivation for the Bonferroni case, which more formally proves Corollary 2.3. Under the alternative $H_{1B}^R(\tau)$, the probability of conflicting test results is zero because all three tests in (7), (8) and (9) are consistent. However, contradictory test results can still occur under the alternative hypothesis in finite sample, which we demonstrate via simulations in Section 3.

A coherent test for jumps in yield spreads takes into account the inherent multiplicity by jointly testing for jumps in the underlying bond yields. A joint test considers different
partitions of the event-return space and motivates the following intersection union test.

**Proposition 2.4** Based on the elementary tests provided in (7), (8) and (9), the intersection union test (IUT) for jumps in yield spreads at time $\tau$ and significance level $\alpha$,

$$\varphi^S_{\alpha}(\text{IUT}, k) = 1 \left\{ \varphi^B_{\alpha}(k) = \varphi^S_{\alpha}(\text{LM}) = 1 \right\}, \quad k = \chi^2, \text{Bonf},$$

(10)

is a coherent test of $H_0^S(\tau)$ with strong error rate control, $P \left( \varphi^S_{\alpha}(\text{IUT}, k) = 1 \right) \leq \alpha$.

In analog to the step-wise procedure of Hsu and Berger (1999), we first verify the more restrictive hypothesis $H^B_0(\tau)$ using either $\varphi^B_{\alpha}(\chi^2)$ or $\varphi^B_{\alpha}(\text{Bonf})$ at level $\alpha$ in the first step. If $H^B_0(\tau)$ is rejected, we proceed to test $H^S_0(\tau)$ at level $\alpha$ with $\varphi^S_{\alpha}(\text{LM})$ in the second step. A jump in the yield spread is detected if and only if the hypotheses in both steps are rejected. The IUT in fact tests the union of both elementary null hypotheses against the intersection of their alternatives. The two-step procedure controls the probability of a false rejection to be $\alpha$ at most for all possible configurations of true null hypotheses without the need of any multiplicity correction across the two steps. It follows the closed testing principle of Marcus et al. (1976) to test hypotheses that are nested in sequence, starting with the most restrictive one. In the current context, the closed testing principle requires rejecting the intersection $H^B_0(\tau) \cap H^S_0(\tau)$ at level $\alpha$ in the first step, before testing $H^S_0(\tau)$ at level $\alpha$ in the second step. The two steps of the closed test procedure are compatible to those of the IUT procedure, because $H^B_0(\tau)$ is a special case of $H^S_0(\tau)$, and $H^B_0(\tau) \cap H^S_0(\tau)$ is identical to the more restrictive hypothesis $H^B_0(\tau)$. The closed test principle provides a link to the power considerations in Finner and Strassburger (2002), see also Goeman and Solari (2010).

We illustrate the IUT using data on the U.S. nominal and inflation-indexed bond yields on October 24, 2019, which are used in constructing Figure 1. Figure 2 depicts the acceptance and rejection regions of the IUT at level $\alpha = 0.01$. The confidence set of the IUT takes the union of the confidence sets in the two steps. Panel (a) combines the square-shaped confidence set of the Bonferroni-adjusted test $\varphi^B_{\alpha}(\text{Bonf})$ on the two bond yields in
Figure 2: Partitions of the event-return space identified by the IUT.

(a) $\varphi_S^\alpha$(IUT, Bonf)

(b) $\varphi_S^\alpha$(IUT, $\chi^2$)

Note: The gray shaded area are the confidence sets of the respective IUT at $\alpha = 0.01$. The white areas form the intersection alternative, where both elementary tests reject their null hypotheses, and a jump in the yield spread is detected. Data used in constructing these plots are the same as in Figure 1, i.e. 30-second observations on U.S. 10-year nominal and inflation-indexed bonds on October 24, 2019.

the first step, with the diagonal corridor, which is the confidence set for the univariate jump test on the spread $\varphi_S^\alpha$(LM) in the second step. Panel (b) uses the $\chi^2$ bivariate jump test $\varphi_B^\alpha(\chi^2)$ in the first step, which replaces the square in Panel (a) by an ellipse. The variance-covariance matrix of the event return $\Gamma_\tau$ determine the shape of the confidence sets in Figure 2. According to (9), the diagonal corridor is narrower (wider) if the estimated bond returns have a larger (smaller) covariance. While the covariance has no effect on the square-shaped confidence set of the Bonferroni-based test (7) in Panel (a), a smaller magnitude of the covariance forms the widely stretched elliptical shape of the $\chi^2$ test (8) in Panel (b) towards a circle.

It is evident from Figure 2 that the confidence set of the $\chi^2$-based IUT occupies a smaller area in the event-return space. Hence, compared with the Bonferroni-adjusted approach, the $\chi^2$-based IUT is more powerful in detecting jumps across combinations of event returns of the two bond yields. Moreover, the cost of achieving coherent test outcomes for the $\chi^2$-based IUT relative to the univariate jump test on the spread $\varphi_S^\alpha$(LM) seems small, as its rejection area in Panel (b) is close to that of the univariate jump test on the spread.
We quantify these differences in finite sample in the simulation study.

Remark 1 Because the second step of the IUT applies the Lee-Mykland test \( \varphi^S_{\alpha}(LM) \) on the yield spread at level \( \alpha \), it directly follows that the joint probability that the test \( \varphi^S_{\alpha}(LM) \) detects a jump in the spread but no jump is detected in the individual bond yields using \( \varphi^B_{\alpha}(k), k = \text{Bonf}, \chi^2 \), is the same as the probability of conflicting test results of the coherent and incoherent tests for spread jump \( \varphi^S_{\alpha}(IUT, k) \) and \( \varphi^S_{\alpha}(LM) \):

\[
P \left( \varphi^S_{\alpha}(LM) = 1, \varphi^S_{\alpha}(IUT, k) = 0 \right) = P \left( \varphi^S_{\alpha}(LM) = 1, \varphi^B_{\alpha}(k) = 0 \right), \quad k = \text{Bonf}, \chi^2.
\]

Remark 1 states that for comparing the coherent IUT procedure with the incoherent Lee-Mykland test, we can focus on situations where the Lee-Mykland test finds a jump in the yield spread, while at the same time, the jump test on the two bond yields, either \( \varphi^B_{\alpha}(\text{Bonf}) \) or \( \varphi^B_{\alpha}(\chi^2) \), does not detect a jump. This is because the second step of the IUT is identical to the univariate Lee-Mykland test \( \varphi^S_{\alpha}(LM) \). We use this strategy in the simulation and the empirical study to evaluate the usefulness of the IUT in eliminating the incoherent test outcomes.

3 Simulation study

The simulation compares the IUT against the Lee-Mykland test applied to the spread. According to Remark 1, we focus on situations where we find no jumps in the two bond yields but a jump in the yield spread. We also investigate the power and size properties of the proposed IUT procedure. As these results are closely linked to the simulation studies in Lee and Mykland (2012), we leave them in Appendix (B).

3.1 Simulation design

The simulation setup emulates that of Lee and Mykland (2012) in a bivariate factor stochastic volatility model. We consider observations of the two bond yields on a normalized,
three-hour interval centered around the event time $\tau$. The two bond yields are specified as

$$ dy_t^{(a)} = \sigma_t dW_t + J_t^{(a)} dN_t^{(a)}, $$

$$ dy_t^{(b)} = \rho \sigma_t dW_t + \sqrt{1-\rho^2} \sigma_t \tilde{W}_t + J_t^{(b)} dN_t^{(b)}, \quad t \in (0, 1], $$

with a Heston-type stochastic volatility

$$ d\sigma_s^2 = 0.0162 \left(0.8465 - \sigma_s^2\right) ds + 0.117 \sigma_s dB_s + J_t^{(\sigma)} dN_t^{(\sigma)}. $$

The continuous component of the yields is driven by independent standard Brownian motions $B_t$, $W_t$ and $\tilde{W}_t$. $\rho$ determines the correlation of the continuous component of the two yields. The parameters in (12) are adopted from Lee and Mykland (2012). With 252 trading days per year, the parameters in (12) imply a yearly volatility of $\sqrt{252 \times 0.8465} = 14.6$ and a half-life of roughly two month ($\left((\log 2)/0.0162 = 42.79\right)$. We follow Aït-Sahalia et al. (2020) in the modelling of jumps. The jump component of the two bond yields consists of a systematic and an idiosyncratic part. The idiosyncratic part has jump size $J_t^{(i)}$ which follows a double exponential distribution,

$$ J_t^{(i)} = \begin{cases} 
\exp(g_t^{(i)}) & \text{with probability } p^{(i)} \\
-\exp(g_t^{(i)}) & \text{with probability } 1-p^{(i)}, 
\end{cases} \quad t \in (0, 1], \quad i = a, b, $$

with parameters $(g_t^{(a)}, g_t^{(b)}, p^{(a)}) = (3\sqrt{1/n}, 3\sqrt{1/n}, 0.5)$. Therefore, the average size of the idiosyncratic jumps is $3\sqrt{1/n}$ in absolute value. The timing of idiosyncratic jumps in yields is determined by independent Poisson processes $N_t^{(i)}$, $i = a, b$, with 126 expected jumps per year. In contrast to the idiosyncratic jumps, we restrict the systematic part of the jump processes to the event time $t = \tau$. We set $J_t^{(a)} = J_t^{(b)} = 0$ under the null hypothesis, and $J_t^{(a)} = 0$ and $J_t^{(b)} = 0.3$ when examining the tests under the alternative hypothesis. The latter leads to a jump size of 0.3 on the yield spread. Appendix (B) contains more results on how different jump sizes affect the size and power of the IUT, where we set the jump size as multiples of the estimation noise.
Jumps in volatility affect both bond yields. Therefore, we model volatility jumps at the time of the systematic jump only, that is \( N_t^{(\sigma)} = 1_{\{t=\tau\}} \). The jump size \( J^{(\sigma)}_\tau \) is exponentially distributed with mean 0.04. According to (6), larger volatility jump at \( \tau \) makes it harder to detect jump in the bond yields and the spread between them. Thus, larger volatility jumps elevate the level of the estimation noise in a similar manner as the microstructure noise. The market microstructure noise is simulated as

\[
\epsilon_j^{(i)} = 0.0861 \Delta y_{t_j}^{(i)} + 0.06 (\Delta y_{t_j}^{(i)} + \Delta y_{t_{j-1}}^{(i)}) + U_j, \quad j = 0, \ldots, n, \quad i = a, b,
\]

where \((U_j)_{0 \leq j \leq n}\) is a sequence of normally distributed random variables with mean 0 and variance \(q^2 = 0.01\). The cross-correlation between the yields \(y_t\) and the noise violates one of our theoretical assumptions. However, simulation results show that this correlation does not affect the finite-sample performance of the test. Notice that if the noise in (14) has variance 0.01, it contributes \(\sqrt{2 \times 0.01} = 0.14\) to each of the observed yield changes such that a jump of size \(J^{(b)}_\tau = 0.3\) is about two-times this standard deviation. Appendix (B) provides details on the estimation of the (co)volatility and noise using the approach of Christensen et al. (2010) and Koike (2016).

### 3.2 IUT versus univariate Lee-Mykland test

The simulation uses \(R = 300,000\) repetitions for each combination of significance level \(\alpha \in [0.01, 0.15]\) and correlation \(\rho \in (-1, 1)\). We start with a sampling frequency of 1-second, which leads to \(n = 10,800\) on a three-hour interval. This mimics the asymptotic results by using a large number of high-frequency observations. Figure 3 depicts the frequency of conflicting test outcomes under the null \(\mathbb{H}_0^B(\tau)\) of no jump in the two bond yields (and the spread) for varying \(\alpha\) and \(\rho\). For given values of \(\alpha\) and \(\rho\), the frequency of conflicting test results is defined as the ratio

\[
\frac{\sum_{i=1}^R 1\{\varphi^{B,i}_{\alpha,i}(k) = 0\} 1\{\varphi^{S,i}_{\alpha,i}(LM) = 1\}}{\sum_{i=1}^R 1\{\varphi^{S,i}_{\alpha,i}(LM) = 1\}}, \quad k = \text{Bonf, } \chi^2,
\]

(15)
which estimates the probability that no jump is detected in the two bond yields, conditional on a detected spread jump using the univariate Lee-Mykland test. According to Remark 1, the numerator in (15) is equivalent to the occurrence of incoherent test outcomes where the IUT detects no jump while the Lee-Mykland test applied to the spread detects a jump. The denominator in (15) estimates the probability of detecting a spread jump, which under \( H_0^B(\tau) \) should be close to \( \alpha \).

We highlight two special cases in Figure 3: \( \rho = 0.8 \) combined with \( \alpha = 1\% \) and 10\%, respectively. When using the \( \chi^2 \) test to detect jumps in the two bond yields, Panel (a) shows that the conditional probability of detecting no jump in bond yields but a jump in the spread is 47\% for \( \alpha = 10\% \) and 54\% for \( \alpha = 1\% \). That is, in the case where there is no jump at event time \( \tau \) and the Lee-Mykland jump test applied to the yield spread detects a jump, roughly half of the false discoveries produce conflicting test outcomes.

Since the Bonferroni jump test is conservative, especially when the correlation is large, the corresponding probabilities shown in Panel (b) of Figure 3 are even larger. With \( \rho = 0.8 \), the conditional probability (15) is 85\% for \( \alpha = 10\% \) and 91\% for \( \alpha = 1\% \).

The surfaces in Figure 3 show that conditional probability of conflicting test results...
increases as the significance level $\alpha$ decreases. For the $\chi^2$ test in Panel (a), the slope is small and remains the same across different values of $\rho$. In contrast, the Bonferroni test in Panel (b) displays an increasing slope when the correlation becomes higher. This is an immediate consequence of the fact that the Bonferroni test does not take into account the covariance of the pre-average returns, which is affected by $\rho$ in the simulation. At the same time, a larger correlation leads to a narrower confidence band for the Lee-Mykland test as shown in (9), and hence smaller detected jumps in the spread. The combination of these two tests therefore has an increasing probability of conflicting test outcomes as $\rho$ increases. The $\chi^2$ test depicted in Panel (a) does not suffer from this problem, because its confidence set adapts as $\rho$ changes. This makes the probability of conflicting test results invariant across different correlations.

We next investigate how smaller sample sizes affect the conditional probability of conflicting test results. Setting the correlation $\rho = 0.8$, Panel (a) of Figure 4 is simulated under the global null hypothesis of no jump, and Panel (b) is simulated under the alternative hypothesis of a spread jump. The $x$-axis shows different sample sizes $n$, with $n = 360$ and $n = 10,800$ corresponding to the 30-second and 1-second sampling schemes, respectively. The $y$-axis shows the frequency of conflicting test outcomes for a given significance level $(\alpha = 0.1$ or $\alpha = 0.01)$ and jump test $(k = \chi^2$ or $k =$Bonf). Panel (a) shows that the conditional probability of conflicting test results under the global null is not affected by smaller sample sizes. Consistent with results shown in Figure 3, the $\chi^2$ bivariate jump test has conditional probabilities of around 50% contradicting the Lee-Mykland test outcome on the spread, while the conditional probabilities for the Bonferroni approach is over 80%.

Panel (b) of Figure 4 imposes a jump on the yield spread. The probability of conflicting test results shown in Panel (b) can be interpreted as the power loss of using the IUT over the Lee-Mykland test on the spread alone. The power loss is the larger in smaller samples. For example, with $\alpha = 0.01$ and 15-second to 30-second sampling frequencies, among all
Figure 4: Probability of incoherent test results for varying sample size $n$.

(a) Null hypothesis: No jump

(b) Alternative hypothesis: Spread jump

Note: The $y$-axis shows the frequency of conflicting test outcomes $P(\varphi^B_\alpha(k) = 0|\varphi^S_\alpha(LM) = 1), k = \chi^2$, Bonf, when the sample size along the x-axis varies. Panel (a) is simulated under the global null $H^B_0(\tau)$ of no jump in the two bond yields, and Panel (b) is simulated under the alternative $H^S_1(\tau)$ of a jump in the yield spread. The jump sizes in Panel (b) are $J^{(a)} = 0$, $J^{(b)} = 0.3$, resulting in a spread jump of size 0.3. Simulations are based on $R = 300,000$ repetitions for each sample size $n$. $q = 0.1$ and $\rho = 0.8$.

jumps that the Lee-Mykland test finds in the yield spread, more than 50% of the times the Bonferroni approach does not find jumps in the two bond yields. The probability of conflicting test results reduces for larger values of $\alpha$, and is much smaller when using the $\chi^2$ jump test. In fact the power loss in using the coherent $\chi^2$-based IUT is at most 9% for $\alpha = 0.01$ and 3% for $\alpha = 0.1$. Dictated by the consistency of the Lee-Mykland test and the IUT, the probability of conflicting test outcomes reduces to 0 as the sampling frequency increases.

The simulations show that conflicting test results are non-negligible under both the null and the alternative hypotheses. This underlines the benefit of coherent tests such as the IUT from Proposition 2.4 which eliminate such contradictions. In addition, the IUT based on the $\chi^2$ jump test comes at a relatively low cost, in that the power loss in small samples vis-à-vis the Lee-Mykland test applied to the spread is small.
4 Empirical evidence

This section demonstrates the usefulness of the proposed spread jump tests using high-frequency data on U.S. government bond yields. We first use two events as examples to illustrate the characteristics of the data and details of the proposed tests. The event study is then extended to all major macroeconomic news releases from 2017 to 2019.

4.1 Data and two illustrative examples

Intra-day high-frequency data on U.S. government bonds are obtained from Refinitiv DataScope Select. The selection procedure and entire list of bonds are given in Appendix (C). For a given release time, we use 30-seconds mid-quotes on an interval spanning from 1.5 hour before to 1.5 hour after the news release time. Pre-averaged returns and (co)variances build on the same methods as in the simulation, and are described in Appendix (B). Instead of using Equation (4) to estimate the event return, we use a jump window of 1.5 minutes:

\[
\Delta \hat{y}_{[\tau_n]} = \hat{y}_{[\tau_n]+1} - \hat{y}_{[\tau_n]-M_n-2}.
\]

This adjustment accounts for cases where the bond yields pick up slightly before the announcement time or display delayed responses.

Figure 5 plots high-frequency break-even inflation rates (top figures) and their underlying bond yields (bottom figures), as two examples. Panel (a) depicts the yield movements around 8:30 am EST on October 24, 2019, which is also used in constructing Figure 1 in the introduction. Panel (b) shows yield movements a few weeks later around 8:30 a.m. EST on December 6, 2019, on which day the U.S. Bureau of Labor Statistics released the monthly Employment Situation report. The report includes key labor market indicators such as the non-farm payroll employment change. Although on both days the yield spread in the top figures displays abrupt movements at 8:30 a.m., the bottom figures suggest that the events are quite different. While in panel (b) both bond yields display a clear discontinuity, in panel (a) such discontinuities are not as obvious.
Figure 5: The 30-seconds bond yields and yield spread data on two event days.

(a) October 24, 2019

(b) December 6, 2019

Note: Yields are in percentage points. The grey shaded area represents the jump window at the 8:30 a.m. EST news release time. The horizontal bars before and after the jump window show the pre-averaged prices used for estimating the yield changes at the announcement time.

Table 1 tabulates some key summary statistics of the local pre-averaged returns at the event time along with the $p$-values of the jump tests. For the announcement at 8:30 a.m. on October 24, 2019, the $p$-values for the univariate Lee-Mykland test on the two bond yields are 0.49 and 0.046. Confirming the observation from Figure 1 in the introduction, with a family-wise error rate of 1% neither the nominal nor the inflation-indexed (real) bond yields display a jump. The more efficient $\chi^2$-jump test confirms the finding of no jump in the two bond yields at 1% level. At the same time, with a $p$-value of 0.004, the Lee-Mykland test applied to the yield spread detects a jump. A stylized pattern of incoherent test results is a positive correlation between the two event returns combined with opposite directional event returns of the two underlying bonds. The statistics shown for December, 6 2019, contrast such conflicting test results. The 10-year nominal bond yield has a significantly
Table 1: Local statistics for jump tests at news release times.

<table>
<thead>
<tr>
<th>Day</th>
<th>Bond</th>
<th>Event time τ</th>
<th>Jump test (p-values)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Return</td>
<td>Variance</td>
<td>Corr.</td>
</tr>
<tr>
<td>Oct 24</td>
<td>nominal</td>
<td>-0.130</td>
<td>0.689</td>
</tr>
<tr>
<td></td>
<td>real</td>
<td>0.497</td>
<td>1.180</td>
</tr>
<tr>
<td></td>
<td>spread</td>
<td>-0.627</td>
<td>0.885</td>
</tr>
<tr>
<td>Dec 6</td>
<td>nominal</td>
<td>6.260</td>
<td>1.611</td>
</tr>
<tr>
<td></td>
<td>real</td>
<td>4.473</td>
<td>1.392</td>
</tr>
<tr>
<td></td>
<td>spread</td>
<td>1.787</td>
<td>0.824</td>
</tr>
</tbody>
</table>

Note: The event returns are the pre-average yield changes in basis points. The event time τ is 8:30 a.m. EST for both days. The correlation (Corr.) of the pre-average event returns of the two bonds is calculated as \( \frac{\Gamma_{[\tau_n]}(i,j)}{\sqrt{\Gamma_{[\tau_n]}(i,i)\Gamma_{[\tau_n]}(j,j)}} \). The sample size \( n = 360 \) for each news release.

larger jump than its inflation-indexed counterpart. Hence, the news release on that day triggers a jump in the 10-year break-even inflation rate, suggesting a significant revision of financial markets’ inflation expectation and the inflation risk premium.

4.2 Conflicting test results

We extend the analysis beyond the two examples from Section 4.1, and implement local jump tests at times of various macroeconomic news releases from January 2017 to December 2019. The complete list of announcements is provided in Appendix (C). There are roughly 500 events that can be classified into four broad categories based on whether it is news on the price level, output, employment, or consumption. In addition to the break-even inflation rates, we also provide empirical evidence on jumps in term spreads.

We conduct local level-α tests at each and every event time and report summary statistics of rejections and no-rejections. That is, our statistics are informative about each specific news release and whether it is effective in moving the bond yields and the spread. The empirical setup does not test any global hypothesis across different news releases. The
Figure 6: Percentage of detected spread jumps without a jump in bond yields.

(a) Break-even inflation

(b) Term spread

Note: Percentage is computed using the ratio (15). Local level-\(\alpha\) tests are conducted at major macroeconomic news release times between 2017 and 2019. Spread jumps are detected using the univariate Lee-Mykland test in (9). Jumps in the two bond yields are tested using the Bonferroni-adjusted approach (7) or the \(\chi^2\) test (8).

testing of global hypotheses require a joint error-rate control across events with potentially very small local level \(\alpha\). Because of limited data deep in the tails of return distributions, an empirical assessment of locally incoherent test results is not feasible for \(\alpha\) much smaller than 1%.\(^3\)

Using the same statistic (15) as in the simulations, Figure 6 illustrates the frequency of incoherent test results occurring. These are the percentage of news events where no jumps are identified in the two underlying bond yields, among all events with a locally detected jump in the corresponding yield spread using the Lee-Mykland test. Panel (a) depicts the percentage of incoherent test results for the break-even inflation rates at three different horizons (5-year, 10-year, and 20-year), and Panel (b) for the three term spreads (2-year-5-year, 2-year-10-year, and 2-year-20-year). Two main results stand out from Figure 6. First, for a given local level \(\alpha\), the Bonferroni-adjusted approach to detect jumps in bond yields always leads to a higher percentage of incoherent test results than the \(\chi^2\) test. For example,

\(^3\) Incoherent test outcomes are a subset of rejections at the rejection boundary. The smaller the \(\alpha\), the more extreme we are in the tails of bivariate return distributions, and the fewer conflicting statistics are detectable. A potential degeneracy is the consequence of data limitation far in the tail, and not because the incoherent test results do not exist for small local \(\alpha\). In fact, the simulation results in Figure 3 demonstrate that incoherent test outcomes become more likely with smaller \(\alpha\).
using the Bonferroni-adjusted test at level $\alpha = 0.05$, 14 of the 90 jump events (15.6%) in the 5-year break-even-inflation rate detected using the univariate Lee-Mykland test do not have corresponding jumps detected in neither of the 5-year nominal and inflation-indexed bond yields. Replacing the Bonferroni approach by the $\chi^2$ test, the percentage of conflicting test results reduces to 6 out of 90 (6.7%) detected spread jumps. The difference between the Bonferroni and $\chi^2$ test results is larger for the break-even inflation rates in Panel (a) than the term spreads in Panel (b). For example, 11 out of 121 (9.1%) detected jumps in the 2-year-5-year term spread are incoherent if we apply the Bonferroni approach at the 10% level, while this number is 10 out of 121 (8.3%) for the $\chi^2$ test at the same level. As shown in Figure 3 in the simulation section, similar probabilities of incoherent test results for the $\chi^2$ and Bonferroni approach at a given test level occur if pre-average event returns display small correlations. In such cases confidence regions of the $\chi^2$ test and Bonferroni test overlap the most. This is exactly what we observe in the data: the returns in two nominal bond yields with different maturities are in general less correlated than the returns in a pair of nominal and inflation-indexed bond yields with the same maturity.

The second main result compares across the different significance levels. For the Bonferroni approach in Figure 6, Panel (a), a smaller test-level comes with a larger frequency of incoherent results. The frequency peaks at 24.2% for the 5-year break-even rate and $\alpha = 0.01$, when 16 out of the 66 jumps in 5-year break-even inflation are not coherent. The downward-sloping pattern for larger $\alpha$ is consistent with the simulation result shown in Figure 3. The slope decreases for the less correlated event returns in the term spreads shown in Panel (b), where the frequency of incoherent test results is more equally distributed than in Panel (a). The lower correlation in the event returns of term spreads than those of break-even inflation is confirmed below in Table 2. In contrast to the Bonferroni-based approach, the percentage of conflicting results using the $\chi^2$ approach remains stable across different $\alpha$. This is in line with the results shown in Figure 3 in the simulation, where the
Table 2: Sizes of event returns for the bond yields and yield spreads.

<table>
<thead>
<tr>
<th></th>
<th>Break-even inflation</th>
<th>Term spread</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BEI5Y</td>
<td>BEI10Y</td>
</tr>
<tr>
<td>Incoherent cases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond 1</td>
<td>0.27</td>
<td>0.23</td>
</tr>
<tr>
<td>bond 2</td>
<td>0.25</td>
<td>0.27</td>
</tr>
<tr>
<td>spread</td>
<td>0.52</td>
<td>0.48</td>
</tr>
<tr>
<td>corr.</td>
<td>0.47</td>
<td>0.71</td>
</tr>
<tr>
<td>Coherent cases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond 1</td>
<td>1.22</td>
<td>1.21</td>
</tr>
<tr>
<td>bond 2</td>
<td>1.07</td>
<td>0.97</td>
</tr>
<tr>
<td>spread</td>
<td>0.65</td>
<td>0.65</td>
</tr>
<tr>
<td>corr.</td>
<td>0.51</td>
<td>0.56</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incoherent cases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond 1</td>
<td>0.22</td>
<td>0.52</td>
</tr>
<tr>
<td>bond 2</td>
<td>0.23</td>
<td>0.20</td>
</tr>
<tr>
<td>spread</td>
<td>0.44</td>
<td>0.64</td>
</tr>
<tr>
<td>corr.</td>
<td>0.48</td>
<td>0.44</td>
</tr>
<tr>
<td>Coherent cases</td>
<td></td>
<td></td>
</tr>
<tr>
<td>bond 1</td>
<td>1.17</td>
<td>1.19</td>
</tr>
<tr>
<td>bond 2</td>
<td>1.02</td>
<td>0.95</td>
</tr>
<tr>
<td>spread</td>
<td>0.64</td>
<td>0.64</td>
</tr>
<tr>
<td>corr.</td>
<td>0.52</td>
<td>0.56</td>
</tr>
</tbody>
</table>

Note: Detected spread jumps are based on a local test level of 5%. Average absolute magnitude of pre-average event return in basis points. Upper panel refers to the Bonferroni approach (7) to detect jumps in bond yields, lower panel to the \( \chi^2 \) approach (8). Coherent cases detect a jump in bond yields and the spread, while incoherent cases do not display a jump in bond yields. Corr. is the correlation of pre-average event returns. Based on macroeconomic data releases between 2017 and 2019.

correlation has no effect on the probability of conflicting test results for the \( \chi^2 \) test.

Table 2 presents the average absolute magnitude of the event returns when a jump in at least one of the two bond yields is detected and the Lee-Mykland test on the spread displays a jump (coherent cases; IUT as defined in Proposition 2.4). This is contrasted with jump sizes when conflicting test results between the Lee-Mykland test on the yield spread and the underlying bond yields occur. There are several common features across the Bonferroni-based approach in the top panel and \( \chi^2 \)-based approach in the bottom panel.
Firstly, when test results are coherent, the average event returns of the two bond yields are larger than the average event return of the yield spread, suggesting jumps in the two bond yields are usually towards the same direction. On the other hand, when test results are incoherent, the average event return of the spread is larger than those of the two bond yields. This indicates that in the incoherent cases, yield changes of the two bonds are usually in opposite directions. Secondly, the average sizes of the event returns of the two bond yields are much larger when test results are coherent compared to when conflicting test results occur. Last but not least, the correlation between the nominal and inflation-indexed bonds of the same maturity is always higher than the correlation between two nominal bonds of different maturities. This confirms the discussion on the percentage of incoherent results shown in Figure 6.

### 4.3 Detected spread jumps

Table 3 summarized the test results using the three tests from Section 2.4 on the break-even inflation rates and term spreads. The three tests are the Lee-Mykland test on the yield spread, and the two IUTs with either the Bonferroni or the $\chi^2$ approach. The first row of each panel reports the number of locally detected jumps across all news releases. As expected from the discussion of the incoherent test results and Remark 1, the univariate Lee-Mykland test applied to the yield spread always finds the highest number of jumps, while the IUT with the Bonferroni approach is the most conservative. Longer maturity break-even inflation and term spreads tend to have more jumps than their shorter horizon counterparts. Thus, longer term bond yields display a more distinctive local response to the release of macroeconomic data.

We further classify the detected jumps in yield spreads according to the four news categories, and calculate the percentage of news releases in each category that leads to a jump in the spread. Although the specific number of detected jumps varies across the three
Table 3: Frequencies of detected jumps in the yield spreads at macroeconomic news release times.

<table>
<thead>
<tr>
<th></th>
<th>Break-even inflation</th>
<th>Term spread</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BEI5Y</td>
<td>BEI10Y</td>
</tr>
<tr>
<td>LM</td>
<td>90</td>
<td>81</td>
</tr>
<tr>
<td>Price</td>
<td>31.6%</td>
<td>27.6%</td>
</tr>
<tr>
<td>Output</td>
<td>14.4%</td>
<td>14.1%</td>
</tr>
<tr>
<td>Employment</td>
<td>18.8%</td>
<td>16.6%</td>
</tr>
<tr>
<td>Consumption</td>
<td>17.0%</td>
<td>19.4%</td>
</tr>
<tr>
<td>IUT(Bonf)</td>
<td>76</td>
<td>76</td>
</tr>
<tr>
<td>Price</td>
<td>27.6%</td>
<td>27.6%</td>
</tr>
<tr>
<td>Output</td>
<td>13.9%</td>
<td>12.6%</td>
</tr>
<tr>
<td>Employment</td>
<td>15.2%</td>
<td>15.7%</td>
</tr>
<tr>
<td>Consumption</td>
<td>14.0%</td>
<td>18.4%</td>
</tr>
<tr>
<td>IUT($\chi^2$)</td>
<td>84</td>
<td>78</td>
</tr>
<tr>
<td>Price</td>
<td>28.9%</td>
<td>27.6%</td>
</tr>
<tr>
<td>Output</td>
<td>14.4%</td>
<td>12.6%</td>
</tr>
<tr>
<td>Employment</td>
<td>17.9%</td>
<td>16.6%</td>
</tr>
<tr>
<td>Consumption</td>
<td>14.0%</td>
<td>19.4%</td>
</tr>
</tbody>
</table>

Note: Detected spread jumps are based on a local test level of 5%. The number given in the first row of each panel breaks down to the four categories of news as reported in Table C.1. LM refers to the Lee-Mykland test for spread jumps (9). IUT(·) refers to the intersection-union tests from Proposition 2.4, based on either the Bonferroni (Bonf) correction or the $\chi^2$ approach.

Different tests, they all reflect that price-related news releases, including data on employment cost, consumer and producer price indices, lead to jumps in the yield spreads most often. That is, the interpretation of what content causes most often a locally significant revision in spreads is robust across the Lee-Mykland test and the two IUTs. For most yield spreads it is not only that news releases about prices cause the most frequent local responses, also the ranking of the other news categories with respect to their percentage is robust across the three implemented spread jump tests. This suggests that the events that produce incoherent test results almost equally distribute across the news categories and do not cluster at one specific economic content.
5 Conclusion

This paper argues that inference about jumps in the yield spread of two bonds requires a joint approach that includes evidence on jumps in the two underlying bond yield processes. This requirement is an immediate consequence of the high-frequency model, where a jump in the spread exists only if at least one of the underlying yield processes has a jump. Ignoring this inherent connection by basing inference only on a univariate jump test applied to the spread tends to overestimate the number of jumps in yield spreads and puts the coherence of test results at risk. We propose an intersection union test (IUT) for jumps in the spread, which explicitly takes into account the joint jump behavior of the two bond yields and their spread. The IUT uses a two-step procedure that includes testing the hypothesis of no jumps in both underlying bond yields against a jump in at least one of the two yield processes. A $\chi^2$ jump test is proposed for this task. We show using simulations and empirical examples that the proposed coherent test is simple to implement and comes at a relatively small cost of fewer jump classifications compared to a univariate (but incoherent) jump test applied to the spread.

Although the methodology proposed in this paper is illustrated in the context of government bonds and yield spreads, it can be directly applied to other high-frequency asset price data. For example, jumps in the value of a portfolio should occur only if the price of at least one of the assets within the portfolio displays a jump. Similarly, stocks that are cross-listed in different stock exchanges are expected to have the same movements under the different listings. We can use the IUT procedure to discover mis-pricing or synergies between them, especially at the time when earnings announcements are released. We leave these for future research.
References


ONLINE APPENDICES

A Assumptions and technical derivations

Assumptions of the underlying model

In this appendix, we are more precise about the underlying semimartingale model which directly translates from Bibinger et al. (2019). The assumptions impose the maximal degree of generality that still allow the estimation of pre-averaged yields (3) and returns (4) in the context of Proposition 2.1. We consider (1) on some filtered probability space \((\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P})\). The jumps \(J_t\) in (1) are split into compensated (small) jumps and finitely many large jumps:

\[
 J_t = \int_0^t \int_{\mathbb{R}^2} \delta(s, z) \mathbb{1}_{\{|\delta(s, z)| \leq 1\}} (\mu - \nu)(ds, dz) + \int_0^t \int_{\mathbb{R}^2} \delta(s, z) \mathbb{1}_{\{|\delta(s, z)| > 1\}} \mu(ds, dz),
\]  

(A.1)

with the jump size function \(\delta\), defined on \(\Omega \times \mathbb{R}_+ \times \mathbb{R}^2\), and the Poisson random measure \(\mu\), which is compensated by \(\nu(ds, dz) = \lambda(dz) \otimes ds\) with a \(\sigma\)-finite measure \(\lambda\). The smoothness of the elements of the drift \(b_t^{(i)}\) and \(\sigma_t^{(i,j)}\), \(i, j = a, b\) of spot squared volatility \(\Sigma_t = \sigma_t \sigma_t'\) is defined by the following assumption:

**Assumption 1** In (1), for assets \(i, j = a, b\), the drift \((b_t^{(i)})_{t \geq 0}\) is a locally bounded process. The volatilities never vanish, \(\inf_{t \in [0,1]} \sigma_t^{(i,i)} > 0\) almost surely. For all \(0 \leq t + s \leq 1, t \geq 0\), some constants \(C_n, \tilde{C}_n > 0\), some \(\beta > 1/2\) and for a sequence of stopping times \(T_n\) increasing to \(\infty\), we have that

\[
 \mathbb{E}\left[ \sigma_{(t+s)\wedge T_n}^{(i,j)} - \sigma_{t\wedge T_n}^{(i,j)} \mid \mathcal{F}_t \right] \leq C_n \, s^\beta, \tag{A.2}
\]
\[
 \mathbb{E}\left[ \sup_{t \in [0,s]} \left| \sigma_{(t+s)\wedge T_n}^{(i,j)} - \sigma_{t\wedge T_n}^{(i,j)} \right|^2 \right] \leq \tilde{C}_n \, s. \tag{A.3}
\]

We impose the following regularity conditions on the (co)jumps
**Assumption 2** Assume for the predictable function $\delta$ in (A.1) that $\sup_{\omega,x} |\delta(t,x)|/\gamma(x)$ is locally bounded with a non-negative deterministic function $\gamma$ that satisfies
\[
\int_{\mathbb{R}^2} (\gamma^r(x) \wedge 1) \lambda(dx) < \infty, \tag{A.4}
\]
with jump activity index $r$, $0 \leq r < 4/3$.

The index $r$ in (A.4) measures the (co)jump activity of the bond yields in (1). Smaller values of $r$ make (A.2) more restrictive. $r = 0$ results in finite-activity jumps and $r = 1$ implies jumps that are summable. The upper bound on $r$ is proved by Bibinger et al. (2019) to make the univariate version of Proposition 2.1 hold.

**Proof of Proposition 2.1**

We fill the missing part of the proof of Proposition 3.1 of Bibinger et al. (2019) for the bivariate model. We state here only the crucial extensions of the covariance of the Brownian component and the noise. The higher order $n$ of the drift part allows us to neglect the drifts. Properties of the pre-averaged estimator (drift, Brownian and jump parts) for the individual bonds $i = a, b$, including the mixed normality is shown in Bibinger et al. (2019), and carry over to the bivariate setting. Hence the missing part which proves Proposition 2.1 is the covariance between the Brownian components $C_i$ and noise $\epsilon$ of the two assets at some known stopping time $\tau$, respectively.

We rewrite the vector of pre-averaged returns of the observed yields in terms of increments $\Delta \tilde{y}_j = \tilde{y}_j - \tilde{y}_{j-1}$, and study the independent Brownian and noise component separately,
\[
M_n^{-1} \left( \sum_{k=0}^{M_n-1} \tilde{y}_{[\tau n]+k} - \sum_{k=-M_n}^{-1} \tilde{y}_{[\tau n]+k} \right) = M_n^{-1} \sum_{k=0}^{M_n-1} \left( \tilde{y}_{[\tau n]+k} - \tilde{y}_{[\tau n]+k-M_n} \right)
= M_n^{-1} \left( \sum_{k=1}^{M_n-1} \Delta \tilde{y}_{[\tau n]+k} (M_n - k) + \sum_{k=0}^{M_n-1} \Delta \tilde{y}_{[\tau n]-k} (M_n - k) \right). \tag{A.5}
\]

The strategy of the proof in Bibinger et al. (2019) is then to exploit the above equation
with respect to the individual signal parts of the process \( y_t^{(i)}, i = a, b \) in (2) and (1). For the covariance of the increments of the Brownian components this gives:

\[
\text{Cov} \left[ \sum_{k=1}^{M_n-1} \Delta C_{(\lfloor \tau_n \rfloor+k)/n} - \frac{M_n-k}{M_n} \sum_{k=0}^{M_n-1} \Delta C_{(\lfloor \tau_n \rfloor-k)/n} - \frac{M_n-k}{M_n} \right] = \sum_{k=1}^{M_n-1} \text{E} \left[ \Delta C_{(\lfloor \tau_n \rfloor+k)/n} \Delta C_{(\lfloor \tau_n \rfloor+k)/n} \right] \left( 1 - \frac{k}{M_n} \right)^2
\]

\[
+ \sum_{k=0}^{M_n-1} \text{E} \left[ \Delta C_{(\lfloor \tau_n \rfloor-k)/n} \Delta C_{(\lfloor \tau_n \rfloor-k)/n} \right] \left( 1 - \frac{k}{M_n} \right)^2
\]

with uncorrelated increments on disjoint intervals in case of stochastic volatility. Itô isometry,

\[
\text{E} \left[ \int_0^t \sigma_s^{(a,a)} dW_s^{(a)} \int_0^t \sigma_s^{(b,b)} dW_s^{(b)} \right] = \int_0^t \text{E} \left[ \sigma_s^{(a,a)} \sigma_s^{(b,b)} \right] \rho_s^{(a,b)} ds,
\]

and the smoothness of the volatility and correlation imply that

\[
\text{E} \left[ \Delta C_{(\lfloor \tau_n \rfloor+k)/n} \Delta C_{(\lfloor \tau_n \rfloor+k)/n} \right] \leq \text{E} \left[ \int_{(\lfloor \tau_n \rfloor+k)/n}^{(\lfloor \tau_n \rfloor+k-1)/n} \sigma_s^{(a,a)} ds \right] \leq \text{O}(n^{-2})
\]

\[
= \frac{\rho_s^{(a,b)} \sigma_s^{(a,a)} \sigma_s^{(b,b)}}{n} + \text{O} \left( \sqrt{\frac{M_n}{n^n}} \right),
\]

for \( k = 1, \ldots, M_n - 1 \). Similarly, we obtain to the left of \( \tau \)

\[
\text{E} \left[ \Delta C_{(\lfloor \tau_n \rfloor-k)/n} \Delta C_{(\lfloor \tau_n \rfloor-k)/n} \right] \leq \text{E} \left[ \int_{(\lfloor \tau_n \rfloor-k)/n}^{(\lfloor \tau_n \rfloor-k-1)/n} \sigma_s^{(a,a)} ds \right] \leq \text{O}(n^{-2})
\]

\[
= \frac{\rho_s^{(a,b)} \sigma_s^{(a,a)} \sigma_s^{(b,b)}}{n} + \text{O} \left( \sqrt{\frac{M_n}{n^n}} \right).
\]

The increments in \text{iid} noise contribute

\[
\text{E} \left[ \Delta \epsilon_{(\lfloor \tau_n \rfloor-k)} \Delta \epsilon_{(\lfloor \tau_n \rfloor-k)} \right] = \text{E} \left[ (\epsilon_{(\lfloor \tau_n \rfloor-k)} - \epsilon_{(\lfloor \tau_n \rfloor-k-1)}) (\epsilon_{(\lfloor \tau_n \rfloor-k)} - \epsilon_{(\lfloor \tau_n \rfloor-k-1)}) \right] = 2\eta^{(a,b)}.
\]

Finally, in conjunction with the identities

\[
\sum_{k=1}^{M_n-1} \left( 1 - \frac{k}{M_n} \right)^2 = \frac{1}{3} M_n - \frac{1}{2} + \frac{1}{6} M_n^{-1}, \quad \text{and} \quad \sum_{k=0}^{M_n-1} \left( 1 - \frac{k}{M_n} \right)^2 = \frac{1}{3} M_n + \frac{1}{2} + \frac{1}{6} M_n^{-1},
\]
we obtain the asymptotic covariance of event returns of asset \( a \) and \( b \):

\[
\sqrt{M_n} \mathbb{E} \left[ \Delta y^{(a)}_{\tau n} \Delta y^{(b)}_{\tau n} \right] \to \left( \frac{\rho^{(a,b)}_\tau \sigma^{(a,a)}_\tau \sigma^{(b,b)}_\tau}{3} + \frac{\rho^{(a,b)}_\tau \sigma^{(a,a)}_\tau \sigma^{(b,b)}_\tau}{3} \right) c^2 + 2\eta^{(a,b)}. \quad (A.6)
\]

The positivity of \( \Gamma_\tau \) is a direct consequence of the additive structure in (6) and the positivity of the noise covariance matrix \( \eta \).

**Proof of Corollary 2.3**

We provide a general analytic expression for the dotted region in Figure 1 that relates to the area in the event-return space where incoherent test results occur. To simplify notation, we consider random variables \( x = n^{1/4} \Delta y^{(a)}_\tau \), \( y = n^{1/4} \Delta y^{(b)}_\tau \). Symmetry allows us to focus on the upper rejection area. Integration boundaries in the \( x \) and \( y \) dimension are determined by the Bonferroni test (7) and the Lee-Mykland test (9). The integration bounds of \( x \) are

\[
\text{lower: } x_1(\alpha, \Gamma_\tau^{(a,a)}) = (\Gamma_\tau^{(a,a)})^{1/2} q_\alpha(N),
\]

\[
\text{upper: } x_2(\alpha, \Gamma_\tau) = (\Gamma_\tau^{(b,b)})^{1/2} q_{1-\alpha/2}(N) - (\Gamma_\tau^{(a,a)} + \Gamma_\tau^{(b,b)} - 2\Gamma_\tau^{(a,b)})^{1/2} q_{1-\alpha}(N).
\]

These are the \( x \) coordinates, where the upper border of the diagonal corridor crosses the square. The corresponding coordinates of \( y \) determine the integration bounds for \( y \):

\[
\text{lower: } y_1(x, \alpha, \Gamma_\tau) = x + (\Gamma_\tau^{(a,a)} + \Gamma_\tau^{(b,b)} - 2\Gamma_\tau^{(a,b)})^{1/2} q_{1-\alpha}(N),
\]

\[
\text{upper: } y_2(\alpha, \Gamma_\tau^{(b,b)}) = (\Gamma_\tau^{(b,b)})^{1/2} q_{1-\alpha/2}(N).
\]

Equipped with those bounds and the bivariate normality result from Proposition 2.1, we can express the joint probability of conflicting test results

\[
P (\varphi^S_\alpha(LM) = 1, \varphi^B_\alpha(Bonf) = 0) = 2 \int_{x_1(\alpha, \Gamma_\tau^{(a,a)})}^{x_2(\alpha, \Gamma_\tau)} \int_{y_1(x, \alpha, \Gamma_\tau)}^{y_2(x, \alpha, \Gamma_\tau)} \phi(x, y, \Gamma_\tau) \, dy \, dx,
\]

where \( \phi(\cdot) \) refers to the bivariate normal distribution function. The probability is positive as soon as upper integration bounds are larger than the lower integration bounds, which is
always true, given the $\alpha$ level of both tests. ■
B Additional simulation results

This section contains the simulation results of the the spread jump tests. We report frequencies of jump detection of the univariate Lee-Mykland jump test applied to the spread (9), as well as the IUT in Proposition 2.4.

We use the same bivariate stochastic volatility model with price and volatility jumps as introduced in Section 3. The (co)volatility is estimated using the pre-averaging method of Christensen et al. (2010) with a window size of $\lceil \sqrt{n} \rceil$. We apply the universal threshold with the median absolute deviation of pre-averaged returns to truncate jumps in the estimation of (co)volatilities (see Koike, 2016, immediately after Theorem 5.1). The market microstructure noise is estimated based on equation (12) of Christensen et al. (2010). We use the ‘yuima’ package in R for our computations.

The pre-average estimator of the event return at $t = \tau$ uses a block size of $M_n = \lceil \sqrt{n}/18 \rceil$. The constant $c = 1/18$ is chosen according to Table 5 of Lee and Mykland (2012). The simulation results do not change much by slightly increasing $M_n$. We simulate jumps at the event time $t = \tau$ whose sizes are multiples of the pre-average estimation noise $\gamma$, defined as

$$\gamma = n^{-1/4}(\Gamma_t^{(i,i)})^{1/2}, \quad i = a, b,$$

(B.1)

with $\Gamma_t^{(i,i)}$ as in (6). Since the estimation noise $\gamma$ directly relates to the asymptotic distribution of the pre-average return estimator, it determines the detection properties of the jump tests. The detection of a jump in yields becomes more difficult if: (i) the noise level $q$ is higher; (ii) the volatility of the Brownian component is larger; and (iii) the sample size $n$ is smaller. However, as we define jump sizes as multiples of $\gamma$, the simulated jump sizes increase in $\gamma$. This allows studying how estimation precision of the pre-average estimators and the noise level affect the test decisions. Notice that the simulation setup provides both bonds ($i = a, b$) with the same integrated volatility and noise level, and hence $\gamma$ does not
depend on the specific bond.

Table B.1 shows the rejection frequencies of the three spread jump tests at level $\alpha = 1\%$. The three different spread jump tests are the univariate Lee-Mykland test applied to the spread alone, and the IUT with either the Bonferroni approach or the $\chi^2$ bivariate jump test in the first step. We consider two different sampling frequencies, 30-second ($n = 360$) and 5-second ($n = 2160$). The top panel of Table B.1 reports the simulation results when the noise variance is low ($q^2 = 0.0001$), and the bottom panel for high noise variance ($q^2 = 0.01$). Each simulation is repeated 3,000 times. The jump sizes reported in columns three and four belong to the null hypothesis of no spread jump, while all other columns correspond to the alternative of a jump in the spread.

Under the null of no spread jump, all three tests exhibit reasonable size properties, with actual sizes below the nominal level of $1\%$. We conduct two experiments: (i) Neither of the two underlying bond yields has a jump at time $\tau$ (column three of Table B.1); (ii) Both bond yields have a jump of the same size (column four of Table B.1). In the latter case, the IUT detects jumps in the bond yields with high probability in the first step, and its test outcome is almost fully determined by the test for equal returns in the second step. As a result, rejection rates of all three tests are almost identical across different noise levels and sample sizes. In particular, there is little difference between the Bonferroni and $\chi^2$-based IUT tests under the null hypothesis.

The Power advantage of the $\chi^2$-based IUT over the Bonferroni approach becomes apparent when the spread jump is induced by a jump in only one of the two bond yields. These results are shown in columns five to seven of Table B.1. The Bonferroni-based IUT always has lower power than the $\chi^2$ approach, because it does not make use of the information on the covariance between the two bond yields. When the jump size is small ($2\gamma$) and noise level is high ($q = 0.1$), the power loss of using the Bonferroni-based IUT compared to the univariate Lee-Mykland test on the spread is well above $50\%$. These results are consistent
### Table B.1: Rejection frequencies of spread jump tests.

<table>
<thead>
<tr>
<th>Jump size</th>
<th>Bond a</th>
<th>0</th>
<th>4γ</th>
<th>2γ</th>
<th>3γ</th>
<th>4γ</th>
<th>4γ</th>
<th>5γ</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bond b</td>
<td>0</td>
<td>4γ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2γ</td>
<td>2γ</td>
<td></td>
</tr>
</tbody>
</table>

Noise level: \( q = 0.01 \)

<table>
<thead>
<tr>
<th>30-sec ((n=360))</th>
<th>LM</th>
<th>0.009</th>
<th>0.009</th>
<th>0.309</th>
<th>0.679</th>
<th>0.926</th>
<th>0.302</th>
<th>0.670</th>
</tr>
</thead>
<tbody>
<tr>
<td>IUT(Bonf)</td>
<td>0.000</td>
<td>0.008</td>
<td>0.044</td>
<td>0.168</td>
<td>0.415</td>
<td>0.168</td>
<td>0.486</td>
<td></td>
</tr>
<tr>
<td>IUT((\chi^2))</td>
<td>0.003</td>
<td>0.009</td>
<td>0.227</td>
<td>0.596</td>
<td>0.896</td>
<td>0.275</td>
<td>0.649</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5-sec ((n=2,160))</th>
<th>LM</th>
<th>0.015</th>
<th>0.011</th>
<th>0.591</th>
<th>0.938</th>
<th>0.998</th>
<th>0.605</th>
<th>0.938</th>
</tr>
</thead>
<tbody>
<tr>
<td>IUT(Bonf)</td>
<td>0.001</td>
<td>0.011</td>
<td>0.136</td>
<td>0.479</td>
<td>0.773</td>
<td>0.506</td>
<td>0.887</td>
<td></td>
</tr>
<tr>
<td>IUT((\chi^2))</td>
<td>0.009</td>
<td>0.011</td>
<td>0.506</td>
<td>0.914</td>
<td>0.996</td>
<td>0.596</td>
<td>0.937</td>
<td></td>
</tr>
</tbody>
</table>

Noise level: \( q = 0.1 \)

<table>
<thead>
<tr>
<th>30-sec ((n=360))</th>
<th>LM</th>
<th>0.000</th>
<th>0.000</th>
<th>0.776</th>
<th>0.999</th>
<th>1.000</th>
<th>0.782</th>
<th>1.000</th>
</tr>
</thead>
<tbody>
<tr>
<td>IUT(Bonf)</td>
<td>0.000</td>
<td>0.000</td>
<td>0.085</td>
<td>0.560</td>
<td>0.960</td>
<td>0.757</td>
<td>0.999</td>
<td></td>
</tr>
<tr>
<td>IUT((\chi^2))</td>
<td>0.000</td>
<td>0.000</td>
<td>0.650</td>
<td>0.999</td>
<td>1.000</td>
<td>0.781</td>
<td>1.000</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5-sec ((n=2,160))</th>
<th>LM</th>
<th>0.001</th>
<th>0.001</th>
<th>0.776</th>
<th>0.998</th>
<th>1.000</th>
<th>0.762</th>
<th>0.998</th>
</tr>
</thead>
<tbody>
<tr>
<td>IUT(Bonf)</td>
<td>0.000</td>
<td>0.001</td>
<td>0.141</td>
<td>0.594</td>
<td>0.941</td>
<td>0.726</td>
<td>0.995</td>
<td></td>
</tr>
<tr>
<td>IUT((\chi^2))</td>
<td>0.000</td>
<td>0.001</td>
<td>0.670</td>
<td>0.995</td>
<td>1.000</td>
<td>0.760</td>
<td>0.998</td>
<td></td>
</tr>
</tbody>
</table>

Note: Jump sizes in the first two rows are given as a multiple of the estimation noise \( \gamma \), which is defined in equation (B.1). Each cell shows the frequency of rejections at significance level \( \alpha = 0.01 \) across 3,000 repetitions. LM indicates the Lee-Mykland test applied to the spread (9). IUT(Bonf) is the Bonferroni-adjusted IUT (7) that uses the Bonferroni-adjusted Lee-Mykland test to test for jumps in the two bond yields in the first step. IUT(\(\chi^2\)) implements the \(\chi^2\) bivariate jump test (8) in the first step.

with findings in Section 3 under the alternative of a spread jump. In contrast, the \(\chi^2\)-based IUT does not suffer from such large power loss over the univariate Lee-Mykland test. The difference in the rejection frequencies between these two tests are below 10%, and approaches 0 as the jump size becomes larger.

The last two columns of Table B.1 show jump detection rates when the jump in the spread is induced by jumps of different sizes in the two bond yields. Compared to situations when the jump in the spread is induced by a jump in only one of the two bond yields, the test power of the univariate Lee-Mykland test is not much affected. This is because it does not take into account the properties of the underlying bond yields. In contrast, the two IUT procedures have higher detection rate. This increase in test power is driven the bivariate jump test in the first step of the IUT, where jumps in the two bond yields are
more easily detected than a jump in only one of them. For the same reason, the difference between the Bonferroni and $\chi^2$-based IUT procedures also become smaller.

Under the alternative hypothesis, it is not surprising that larger sample size $n$ almost always leads to higher detection rate. Its effect is more evident when the noise level is low ($q = 0.01$). Comparing the top and bottom panels of Table B.1, we see that the power of all tests increases when the noise level is higher, keeping other parameters fixed. This is because the simulated jump sizes are multiples of the estimation noise $\gamma$ defined in (B.1), which increases with the variance of the microstructure noise. As a result, the simulated jumps has larger magnitudes in the bottom panel for higher noise level.
C Data

Table C.1: Macroeconomic news releases examined in the empirical analyses.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Category</th>
<th>Frequency</th>
<th>Release time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Consumer price index</td>
<td>Price</td>
<td>Monthly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>Producer price index</td>
<td>Price</td>
<td>Monthly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>Employment cost index</td>
<td>Price</td>
<td>Quarterly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>Gross domestic product</td>
<td>Output</td>
<td>Quarterly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>Durable goods orders</td>
<td>Output</td>
<td>Monthly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>ISM manufacturing</td>
<td>Output</td>
<td>Monthly</td>
<td>10:00 am</td>
</tr>
<tr>
<td>Chicago PMI</td>
<td>Output</td>
<td>Monthly</td>
<td>9:45 am</td>
</tr>
<tr>
<td>Empire state manufacturing</td>
<td>Output</td>
<td>Monthly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>Business inventories</td>
<td>Output</td>
<td>Monthly</td>
<td>10:00 am</td>
</tr>
<tr>
<td>Production and utilization</td>
<td>Output</td>
<td>Monthly</td>
<td>9:15 am</td>
</tr>
<tr>
<td>Employment report</td>
<td>Employment</td>
<td>Monthly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>ADP employment change</td>
<td>Employment</td>
<td>Monthly</td>
<td>8:15 am</td>
</tr>
<tr>
<td>Initial jobless claims</td>
<td>Employment</td>
<td>Weekly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>Personal spending</td>
<td>Consumption</td>
<td>Monthly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>Advance retail sales</td>
<td>Consumption</td>
<td>Monthly</td>
<td>8:30 am</td>
</tr>
<tr>
<td>Consumer confidence</td>
<td>Consumption</td>
<td>Monthly</td>
<td>10:00 am</td>
</tr>
</tbody>
</table>

This section provides information on the U.S. macroeconomic news announcements and Treasury bonds data used in the empirical analyses. Table C.1 presents the list of macroeconomic news announcements we use to investigate jumps in bond yields and yield spreads. These announcements are classified into four broad categories: price, output, employment, and consumption.

The high-frequency data on U.S. Treasury bond yields are obtained from Refinitiv DataScope Select provided by Thomson Reuters Tick History. Tables C.2 provides information on the individual nominal (left column) and inflation-indexed bonds (right column). We use maturities that are closest to 2, 5, 10, and 20 years at the time of each news release. When there are several bonds available, we select the bond that has the highest number of non-zero 30-second returns on the day of the announcement, which is considered to be the
Table C.2: The list of U.S. Treasury bonds and TIPS used in the empirical analyses.

<table>
<thead>
<tr>
<th>Treasury bonds</th>
<th>CUSIP</th>
<th>Coupon</th>
<th>Maturity</th>
<th>TIPS</th>
<th>CUSIP</th>
<th>Coupon</th>
<th>Maturity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>912810ED6</td>
<td>8.125</td>
<td>15/08/2019</td>
<td>912828JX9</td>
<td>2.125</td>
<td>15/01/2019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EM6</td>
<td>7.250</td>
<td>15/08/2022</td>
<td>912828TE0</td>
<td>0.125</td>
<td>15/07/2022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EY0</td>
<td>6.500</td>
<td>15/11/2026</td>
<td>912828S50</td>
<td>0.125</td>
<td>15/07/2026</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810PU6</td>
<td>5.000</td>
<td>15/05/2037</td>
<td>912810QP6</td>
<td>2.125</td>
<td>15/02/2041</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EZ7</td>
<td>6.625</td>
<td>15/02/2027</td>
<td>912810QF8</td>
<td>2.125</td>
<td>15/02/2040</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810PT9</td>
<td>4.750</td>
<td>15/02/2037</td>
<td>912828JE1</td>
<td>1.375</td>
<td>15/07/2018</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810FT0</td>
<td>4.500</td>
<td>15/02/2036</td>
<td>912828SA9</td>
<td>0.125</td>
<td>15/01/2022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810FA1</td>
<td>6.375</td>
<td>15/08/2027</td>
<td>912828QV5</td>
<td>0.625</td>
<td>15/07/2021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EK0</td>
<td>8.125</td>
<td>15/08/2021</td>
<td>912828LA6</td>
<td>1.875</td>
<td>15/07/2019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EN4</td>
<td>7.725</td>
<td>15/11/2022</td>
<td>912810PS1</td>
<td>2.375</td>
<td>15/01/2027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810PW2</td>
<td>4.375</td>
<td>15/02/2038</td>
<td>912828X39</td>
<td>0.125</td>
<td>15/04/2022</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EC8</td>
<td>8.875</td>
<td>15/02/2019</td>
<td>912828V49</td>
<td>0.375</td>
<td>15/01/2027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EE4</td>
<td>8.500</td>
<td>15/02/2020</td>
<td>912828C99</td>
<td>0.125</td>
<td>15/04/2019</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EL8</td>
<td>8.000</td>
<td>15/11/2021</td>
<td>912828UH1</td>
<td>0.125</td>
<td>15/01/2023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810FB9</td>
<td>6.125</td>
<td>15/11/2027</td>
<td>912828MF4</td>
<td>1.375</td>
<td>15/01/2020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EP9</td>
<td>7.125</td>
<td>15/02/2023</td>
<td>912810PV4</td>
<td>1.750</td>
<td>15/01/2028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810E3</td>
<td>5.500</td>
<td>15/08/2028</td>
<td>912828L3</td>
<td>0.375</td>
<td>15/07/2027</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810PX0</td>
<td>4.500</td>
<td>15/05/2038</td>
<td>912810FQ6</td>
<td>3.375</td>
<td>15/04/2032</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EF1</td>
<td>8.750</td>
<td>15/05/2020</td>
<td>912828VM9</td>
<td>0.375</td>
<td>15/07/2023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EQ7</td>
<td>6.250</td>
<td>15/08/2023</td>
<td>912810FD5</td>
<td>3.625</td>
<td>15/04/2028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EG9</td>
<td>8.750</td>
<td>15/08/2020</td>
<td>912828K33</td>
<td>1.375</td>
<td>15/04/2020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810QA9</td>
<td>3.500</td>
<td>15/02/2039</td>
<td>912828NM8</td>
<td>1.250</td>
<td>15/07/2020</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810FF0</td>
<td>5.250</td>
<td>15/11/2028</td>
<td>912810PZ5</td>
<td>2.500</td>
<td>15/01/2029</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810ES3</td>
<td>7.500</td>
<td>15/11/2024</td>
<td>912828R9</td>
<td>0.500</td>
<td>15/01/2028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EH7</td>
<td>7.875</td>
<td>15/02/2021</td>
<td>912828H0</td>
<td>0.625</td>
<td>15/04/2023</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810QB7</td>
<td>4.250</td>
<td>15/05/2039</td>
<td>912828B25</td>
<td>0.625</td>
<td>15/01/2024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810FG8</td>
<td>5.250</td>
<td>15/02/2029</td>
<td>912828PP9</td>
<td>1.125</td>
<td>15/01/2021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810EJ3</td>
<td>8.125</td>
<td>15/02/2051</td>
<td>912828Y38</td>
<td>0.750</td>
<td>15/07/2028</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810FJ2</td>
<td>6.125</td>
<td>15/08/2029</td>
<td>912828WU0</td>
<td>0.125</td>
<td>15/07/2024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810ET1</td>
<td>7.625</td>
<td>15/02/2025</td>
<td>912810FH6</td>
<td>3.875</td>
<td>15/04/2029</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810QC5</td>
<td>4.500</td>
<td>15/08/2039</td>
<td>912828Q60</td>
<td>0.125</td>
<td>15/04/2021</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810QD3</td>
<td>4.375</td>
<td>15/11/2039</td>
<td>912828N5</td>
<td>0.500</td>
<td>15/04/2024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810FM5</td>
<td>6.250</td>
<td>15/05/2030</td>
<td>912810FR4</td>
<td>2.375</td>
<td>15/01/2025</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810QE1</td>
<td>4.625</td>
<td>15/02/2040</td>
<td>912828YL8</td>
<td>0.125</td>
<td>15/10/2024</td>
<td></td>
</tr>
<tr>
<td></td>
<td>912810FP8</td>
<td>5.375</td>
<td>15/02/2031</td>
<td>912828XL9</td>
<td>0.375</td>
<td>15/07/2025</td>
<td></td>
</tr>
</tbody>
</table>
most liquid bond at a given maturity.

Table C.3 presents some statistics of the nominal and inflation-indexed bond yields used in the empirical analyses. These statistics are calculated using 30-second bond yields data on 430 unique announcement dates from 7am to 5pm. Panel A summarizes some simple descriptive statistics of the 30-second returns. The first row reports that the average number of non-zero returns ranges from 230 to 450 for different bond types and maturities. In general, long-dated bonds tend to have more movements in the yields, which partly reflect the liquidity level of the bond. The average yield change is always very close to zero. We also present the mean of the yield changes after taking the absolute value of the change to show the typical size of a 30-second return. It varies between 0.03 to 0.05 basis points for different types of bonds and maturities on these announcement days. Lastly, the standard deviation of the yield changes ranges between 0.09 to 0.18 basis points, which is much larger than the average size of the return.

Table C.3: Descriptive statistics of government bond yields and yield spreads.

<table>
<thead>
<tr>
<th>Maturity</th>
<th>Nominal bonds</th>
<th>Indexed bonds</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2Y</td>
<td>5Y</td>
</tr>
<tr>
<td>Panel A: observed 30-second returns</td>
<td></td>
<td></td>
</tr>
<tr>
<td># $\Delta \tilde{y}_i \neq 0$</td>
<td>269</td>
<td>252</td>
</tr>
<tr>
<td>Mean $\Delta \tilde{y}_i$</td>
<td>-0.0001</td>
<td>-0.0004</td>
</tr>
<tr>
<td>Mean $</td>
<td>\Delta \tilde{y}_i</td>
<td>$</td>
</tr>
<tr>
<td>St.dev. $\Delta \tilde{y}_i$</td>
<td>0.1114</td>
<td>0.1809</td>
</tr>
<tr>
<td>Panel B: microstructure noise</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$p$-value</td>
<td>0.027</td>
<td>0.149</td>
</tr>
<tr>
<td>Rejection rate</td>
<td>93.7%</td>
<td>64.2%</td>
</tr>
<tr>
<td>Noise level</td>
<td>0.105</td>
<td>0.103</td>
</tr>
</tbody>
</table>

Note: The bond data are 30-second observations from 7am to 5pm on 430 macroeconomic news release dates from 2017 to 2019. The reported sizes of the return are in basis points. The $p$-value refers to the autocorrelation based test for microstructure noise proposed by Aït-Sahalia and Xiu (2019). The fraction of rejecting the null of no noise using the same test at 5% significance level is reported in row labeled rejection rate. The average noise level is estimated using Proposition 1 of Lee and Mykland (2012).
Panel B of Table C.3 provides evidence on the prevalence of market microstructure noise in the bond data. We report the median $p$-value and the percentage of rejections at a 5% level for the autocorrelation based noise test proposed by Aït-Sahalia and Xiu (2019). The small $p$-values and high rejection rates of no noise, particularly for shorter-term bond yields and yield spreads, support the importance of our noise-robust method proposed in Section 2. The noise level in the last row of Table C.3 reports the average value of the estimated $\eta$, obtained using the noise estimator of Lee and Mykland (2012). The noise level has similar magnitudes to the high noise level used in the simulation in Section 3 across different bonds and spreads, and is comparable to the standard deviation of the 30-second returns.