Fixed and Long Time Span Jump Tests: New Monte Carlo and Empirical Evidence[*]

Mingmian Cheng[1] and Norman R. Swanson[2]

[1]Sun Yat-sen University and [2]Rutgers University

February 2019

Abstract

Numerous tests designed to detect realized jumps over a fixed time span have been proposed and extensively studied in the financial econometrics literature. These tests differ from "long time span tests" that detect jumps by examining the magnitude of the jump intensity parameter in the data generating process, and which are consistent. In this paper, long span jump tests are compared and contrasted with a variety of fixed span jump tests in a series of Monte Carlo experiments. It is found that both the long time span *CSS* tests of Corradi et al. (2018) and the fixed span tests of Aït-Sahalia and Jacod (2009) exhibit reasonably good finite sample properties, for time spans both short and long. Various other tests suffer from finite sample distortions, both under sequential testing and under long time spans. The latter finding is new, and confirms the "pitfall" discussed in Huang and Tauchen (2005), of using asymptotic approximations associated with finite time span tests in order to study long time spans of data. An empirical analysis is carried out to investigate the implications of these findings, and "time-span robust" tests indicate that the prevalence of jumps is not as universal as might be expected.

Keywords: Jump test, Jump intensity, Sequential testing bias, Fixed time span, Long time span, High-frequency data

JEL classification: C12, C22, C52, C58

---

1 Introduction

In this paper, we add to the financial econometrics literature by carrying out an extensive Monte Carlo and empirical analysis comparing different types of jump tests used in the specification process associated with fitting continuous time models of financial variables.[1] We focus on tests due to Barndorff-Nielsen and Shephard (2006), Lee and Mykland (2008), Aït-Sahalia and Jacod (2009), Corsi et al. (2010), and Podolskij and Ziggel (2010), who study "fixed time span" jump tests; and tests due to Corradi et al. (2014) and Corradi et al. (2018) who study so-called "long time span" jump tests.[2] The reason why a "horse race" comparing alternative jump tests of these varieties is of interest is because it is well known that tests constructed using observed sample paths of asset returns on a "fixed time span", such as a day or a week, are not consistent against non-zero jump intensity, and are sensitive to sequential testing bias. On the other hand, the CSS and $CSS1$ tests that we examine are based on direct evaluation of the data generating process, and are thus consistent and asymptotically correctly sized when the time span, $T \to \infty$, and the sampling interval, $\Delta \to 0$. Put differently, one reason why detecting jumps using long time span tests is of potential interest is that empirical researchers often estimate $DGPs$ after testing for jumps using fixed time span tests. This is problematic if the jump intensity is identically zero, since parameters characterizing jump size are unidentified, and consistent estimation of the rest of the parameters is thus no longer feasible (see Andrews and Cheng (2012)). Moreover, if researchers detect jumps on a particular sample path (say, when using a daily fixed time span jump test), they might conclude that the jump intensity is non-zero. However, if no jumps are found in a sample path, this does not mean there are no jumps in other sample paths, and hence that a DGP should be estimated with no jump component. [3]

The CSS and $CSS1$ tests examined in this paper are based on realized third moments, or tricity, and as discussed above, utilize observations over an increasing time span. Although various tricity-type tests have already been examined in the literature, it is worth noting that only the CSS and $CSS1$ tests are developed using both in-fill and long-span asymptotics. As discussed above,

---

[1] In risk management and financial engineering, investors and researchers often require knowledge of the data generating process (DGP) that governs asset price movements. For example, asset prices are frequently modeled as continuous-time processes, such as (Itô-)semimartingales (see, e.g. Aït-Sahalia (2002a,b), Chernov et al. (2003), and Andersen et al. (2007b)). At the same time, investors and researchers are also interested in nonparametrically estimable quantities such as spot/integrated volatility (see, e.g. Barndorff-Nielsen (2002), Barndorff-Neilsen and Shephard (2003), Todorov and Tauchen (2011), and Patton and Sheppard (2015)), jump variation (see, e.g. Barndorff-Nielsen and Shephard (2004), Andersen et al. (2007a), and Corsi et al. (2009)), leverage effect (see, e.g. Kalnina and Xiu (2017) and Aït-Sahalia et al. (2017)), and jump activity (see e.g. Aït-Sahalia and Jacod (2011) and Todorov (2015)).

[2] Specifically, when focussing on long time span tests, we evaluate the CSS test of Corradi et al. (2018), and the $CSS1$ test of Corradi et al. (2014).

[3] In an interesting paper related to this paper, Huang and Tauchen (2005) discuss issues associated with applying asymptotic approximations used in fixed time span jump tests over an (long time span) entire sample. They suggest that an appropriate way to solve both inconsistency and size distortion problems assoiated with fixed time span jump tests is to use test statistics that are asymptotically valid under a double asymptotic scheme where both $T \to \infty$ and $\Delta \to 0$. This is the approach taken by Corradi et al. (2018).

the use of long-span asymptotics ensures that these tests have power against non-zero intensity in the DGP rather than against realized jumps on a particular sample path with a fixed time span. A key difference between the CSS and $CSS1$ tests is that the latter test sacrifices power by using a rescaled bootstrap to ensure robustness against leverage, while the former test uses thresholding and requires the use of time span, T^+, where $T^+/T \to \infty$, to eliminate the leverage effect. In our experiments, we consider two types of the $CSS1$ jump test (i.e. $CSS1$ and $\widetilde{CSS1}$ tests). Both of these build on earlier work of Aït-Sahalia (2002a,b), and are special cases of the CSS test introduced in Corradi et al. (2018). One test assumes no leverage. The other test is robust to leverage, and is a rescaled version of the no leverage test. Both tests are derived under the assumption that $\mathrm{E}\left(\left(Y_{k\Delta} - Y_{(k-1)\Delta}\right)^3\right) = 0$, whenever there are no jumps, where $Y_{k\Delta} = \ln X_{k\Delta} - \frac{\Delta}{T}\sum_{k=2}^{n}\ln X_{k\Delta}$ and $\ln X_{k\Delta}$ is the kth asset return over a shrinking sampling interval Δ.

In our Monte Carlo analysis, the finite sample properties of three fixed time span tests, as well as the three CSS type tests (i.e. $CSS1$, $\widetilde{CSS1}$ and CSS tests) are investigated.[4] The three "fixed span" tests include the higher order power variation test of Aït-Sahalia and Jacod (2009) (ASJ), the classic bipower variation test of Barndorff-Nielsen and Shephard (2006) (BNS), and the truncated power variation test of Podolskij and Ziggel (2010) (PZ).[5] Our findings can be summarized as follows. First, we show that the finite sample power of daily jump tests against non-zero jump intensity is low, particularly when jumps are infrequent or jump magnitudes are "weak". For instance, when the jump intensity is 0.4 and the jump size parameter is our largest, rejection rates of the ASJ, BNS and PZ tests at a 0.05 significance level are only around 0.26, 0.38, and 0.36, respectively. Second, sequential testing bias grows rapidly as the time span increases. The size of a joint test based on the strategy of sequentially performing many fixed-T daily tests approaches unity very quickly. Even for the most conservative test (i.e., the ASJ test), empirical size is over 0.95 after 50 days. Importantly, we also show that the empirical sizes of fixed-T jump tests over samples with growing time spans also increase in T. However, size distortion accumulates more slowly when using the ASJ test than when using the BNS and PZ tests. Moreover, the power of ASJ test is very good for all long time spans, as long as jumps are not too rare and too weak. Finally, long time span jump tests are adequately sized, as long as T and Δ are carefully chosen, regardless of the presence of leverage. Additionally, power is reasonably good, even when jumps are infrequent and weak, and increases with an increasing time span, T, for fixed Δ.

In our empirical analysis, we examine 5-minute intraday observations between 2006 and 2013 on twelve individual stocks, nine sector ETFs, and one market (SPDR S&P 500) ETF. Our main

---

[4]It is important to note that these tests have a different null hypotheses than fixed time span tests. However, our objective in this paper is not only to examine the finite sample properties of bopth types of tests, but also to compare and contrast the two classes of tests, since both are often used as pre-tests, prior to specifying and estimating $DGPs$ involving jumps.

[5]For a detailed comparison of more fixed span tests, refer to Theodosiou and Zikes (2011) and Dumitru and Urga (2012). These authors concisely summarize and compare a large group of existing jump tests via extensive Monte Carlo experiments.

empirical findings are summarized as follows. First, using daily *ASJ*, *BNS* and *PZ* tests, jumps are widely detected in asset prices over almost all time periods considered. Moreover, in some cases, the annual percentage of jump days seems inconceivably large. For instance, all three tests detect jumps on around 35%–40% of the days in 2006 for two of the ETFs that we examine. Second, these jump percentages have diminished over time. Third, long span jump tests tell a different story. Namely, the *ASJ* test, the $\widetilde{CSS1}$ test, and the *CSS* test indicate far fewer jumps than are found when using daily tests.[6] This finding has important implications for both specification and estimation of asset price models.

The rest of the paper is organized as follows. Section 2 outlines the theoretical framework and introduces notation. Section 3 discusses statistical issues associated with testing for jumps. Section 4 discusses the long time span jump tests that we examine, and Section 5 briefly discusses the extant fixed time span tests examined in the sequel. Section 6 reports results from our Monte Carlo experiments. Section 7 presents the results of our empirical analysis of various individual stock and stock index data. Finally, Section 8 contains concluding remarks.

2   Setup

We use the setup of Corradi et al. (2018). Namely, assume that asset (log-)prices are recorded at an equally spaced discrete interval, $\Delta = \frac{1}{m}$, where m is the total number of observations on each trading day. In our model, we assume that $\Delta \to 0$; or equivalently that $m \to \infty$. Log-prices follow a jump diffusion model defined on some filtered probability space $(\Omega, \mathbb{F}, (\mathcal{F}_t)_{t\geq 0}, \mathbb{P})$, with

$$\mathrm{d}\ln X_t = \mu\mathrm{d}t + \sqrt{V_t}\mathrm{d}W_{1,t} + Z_t\mathrm{d}N_t, \tag{1}$$

where μ is the drift term, $\sqrt{V_t}$ is the spot volatility, and $W_{1,t}$ is an adapted standard Brownian motion (i.e., it is $\mathcal{F}_t$-measurable for each $t \geq 0$). Here, V_t is defined according to either (i), (ii), (iii), or (iv), as follows:

(i) a constant:
$$V_t = v \text{ for all } t; \tag{2}$$

(ii) a measurable function of the state variable:

$$V_t \text{ is } X_t\text{-measurable}; \tag{3}$$

(iii) a stochastic volatility process without leverage:

$$\mathrm{d}V_t = \mu_{V,t}(\theta)\mathrm{d}t + g\left(V_t, \theta\right)\mathrm{d}W_{2,t}, \ \mathrm{E}\left(W_{1,t}W_{2,t}\right) = 0; \tag{4}$$

---

[6]The $\widetilde{CSS1}$ test is a "leverage robust" variant of the $CSS1$ test, and is discussed in detail in Section 4.

where the vector θ parameterizes $\mu_V(\cdot)$ and $g(\cdot)$.

(iv) a stochastic volatility process with leverage:

$$dV_t = \mu_{V,t}(\theta)dt + g(V_t, \theta)\, dW_{2,t},\ \mathrm{E}(W_{1,t}W_{2,t}) = \rho \neq 0. \tag{5}$$

Evidently, the volatility process is quite general, although we do not consider jumps in volatility.

Now, define,

$$\Pr(N_{t+\Delta} - N_t = 1|\mathcal{F}_t) = \lambda_t \Delta + o(\Delta), \tag{6}$$

$$\Pr(N_{t+\Delta} - N_t = 0|\mathcal{F}_t) = 1 - \lambda_t \Delta + o(\Delta), \tag{7}$$

and

$$\Pr(N_{t+\Delta} - N_t > 1|\mathcal{F}_t) = o(\Delta), \tag{8}$$

where λ_t characterizes the jump intensity. The jump size, Z_t, is identically and independently distributed with density $f(z;\gamma)$, where γ parameterizes the jump size distribution. Equation (8) implies that we rule out infinite-activity jumps.

When constructing the fixed time span realized jump tests discussed in the sequel, we remain agnostic about the jump generating process. However, for the case of our long time span jump intensity tests, we must provide a moderate amount of additional structure. This is one of the key trade-offs associated with using either variant of test. In particular, following Corradi et al. (2018), we consider two cases. First, N_t, which is the number of jump arrivals up to t, follows a counting process, such as the widely used Poisson process. In this case, $\lambda_t = \lambda$, for all t. Second, jumps may be "self-exciting", in the sense that the jump intensity follows Hawkes diffusion (see Bowsher (2007) and Aït-Sahalia et al. (2015)), with

$$d\lambda_t = a(\lambda_\infty - \lambda_t)\, dt + \beta dN_s, \tag{9}$$

where $\lambda_\infty \geq 0$, $\beta \geq 0$, $a > 0$, and $a > \beta$, so that the process is mean reverting with $\mathrm{E}(\lambda_t) = \frac{a\lambda_\infty}{a-\beta}$. As noted in Corradi et al. (2018), if $\lambda_\infty = 0$, then $\mathrm{E}(\lambda_t) = 0$, which implies that $\lambda_t = 0$, a.s. for all t. This implies that $N_t = 0$, a.s., for all t. As a result, β, a and γ in this case are all unidentified. On the other hand, if $\lambda_\infty > 0$, then β and γ are identified. But if $\beta = 0$, a is not identified. These observations highlight the importance of pretesting for $\lambda_\infty = 0$ against $\lambda_\infty > 0$, in order to obtain consistent estimation of parameters in the case of Hawkes diffusions.

In light of the above discussion, we are interested in testing

$$H_0 : \lambda = 0$$

versus

$$H_A : \lambda > 0,$$

where λ is the constant jump intensity, in the case of Poisson-type jumps; and is the expectation of the stochastic jump intensity (i.e. $\lambda = \mathrm{E}(\lambda_t)$), in the case of self-exciting jumps.[7] This is a nonstandard inference problem because, under H_0, some parameters are not identified, and a parameter lies on the boundary of the null parameter space.

Before discussing the tests that are compared in our Monte Carlo experiments, we first provide some heuristic motivation for long time span jump testing. This discussion follows Corradi et al. (2014).

3 Heuristic Discussion

In recent years, a large variety of tests for realized jumps have been proposed and studied. One common feature of the preponderance of these tests is that they are all carried out using high-frequency observations over a fixed time span and are justified by in-fill asymptotic theorems. Therefore, they have power against realized jumps over fixed time spans, and none are consistent against the alternative of $\lambda > 0$. This inconsistency issue has also been pointed out and illustrated by Huang and Tauchen (2005) and Aït-Sahalia and Jacod (2009), among others. Many of the fixed time span tests can be considered as Hausman-type tests in which a comparison of two realized measures of the integrated volatility is made, where one is robust to jumps, and one is not. Under the null of no jumps, both consistently estimate the integrated volatility. Under the alternative of jumps, however, the consistency of the non-robust realized measure fails. Instead, it estimates the total quadratic variation that contains the contribution from jump components. As a result, these two realized measures differ in the presence of jumps. In general, Hausman-type tests are designed to detect whether $\sum_{j=N_t}^{N_{t+1}} c_j^2 = 0$ or $\sum_{j=N_t}^{N_{t+1}} c_j^2 > 0$, where N_t denotes the number of jumps up to time t, and c_j is the (random) size of the jumps. However, $\lambda > 0$ does not imply that $\sum_{j=N_t}^{N_{t+1}} c_j^2 > 0$, given that $\Pr(N_{t+1} - N_t > 0) < 1$. Therefore, such tests have power against realized jumps, but not necessarily against positive jump intensity.

Two techniques are often employed in practice to construct the jump-robust realized measures. The first uses multipower variations, such as bipower variation or tripower variation. Under these measures, the effect of jumps is asymptotically "removed" by using the product of consecutive high-frequency observations. The second uses jump thresholding that allows for the separation of jump and continuous components, based on the difference between their orders of magnitude. (see, e.g. Mancini (2009) and Corsi et al. (2010)). Recent higher order power variation tests are motivated by the fact that for $p > 2$, $\sum_{i=1}^{n-1} \left| \ln X_{(t+(i+1)\Delta} - \ln X_{t+i\Delta} \right|^p$ converges to $\sum_{t \leq s \leq t+1} \left| \ln X_s - \ln X_{s-} \right|^p$,

---

[7]Note that $\lambda_\infty = (>) \, 0$ if and only if $\mathrm{E}(\lambda_t) = (>) \, 0$.

where $\sum_{t \leq s \leq t+1} |\ln X_s - \ln X_{s-}|^p$ is strictly positive if there are jumps, and zero otherwise (see, e.g. Aït-Sahalia and Jacod (2009) and Aït-Sahalia et al. (2012)). In this case, however, test power obtains still because of realized jumps, and not because of positive jump intensity.

Additionally, other recent tests related to those discussed above have been proposed that are based on comparisons using pre-averaged volatility measures, in order to obtain tests that are robust to microstructure noise (see, e.g. Podolskij and Vetter (2009a,b) and Aït-Sahalia et al. (2012)).

In the Monte Carlo and empirical experiments reported in this paper, we consider three fixed time span tests based on multipower variation, jump thresholding and higher order power variation, respectively.

Generally, jump tests performed over a fixed time span are designed to distinguish between:

$$\Omega_{t,l}^C = \{\omega : s \to \ln X_s(\omega) \mid \Delta \ln X_s = \ln X_s(\omega) - \ln X_{s-}(\omega) = 0, \ \forall s \in [t, t+l)\}$$
$$\text{and}$$
$$\Omega_{t,l}^J = \{\omega : s \to \ln X_s(\omega) \mid \Delta \ln X_s = \ln X_s(\omega) - \ln X_{s-}(\omega) \neq 0, \ \forall s \in [t, t+l)\},$$

where l indicates a fixed time span. Hence, all of the tests discussed above are dependent upon pathwise behavior. Clearly, one might decide in favor of $\Omega_{t,l}^C$, even if $\lambda > 0$, simply because jumps are by coincidence absent over the interval $[t, t+l)$. Now, in order to carry out a consistent test against positive jump intensity, two approaches may be used. First, one may consider the following joint hypothesis:

$$\Omega_T^C = \cap_{t=0}^{T-1} \Omega_{t,l}^C, \quad \text{as } T \to \infty,$$

versus its negation. Here, the objective is to test the joint null hypothesis that none of the fixed-span sample paths contain jumps. In fact, under mild conditions on the degree of heterogeneity of the process, failure to reject $\Omega_\infty^C = \lim_{T \to \infty} \cap_{t=0}^{T-1} \Omega_{t,l}^C$ implies failure to reject $\lambda = 0$. The difficulty lies in how to implement a test for Ω_T^C, when T gets large. Needless to say, sequential application of fixed time span jump tests leads to sequential test bias, and for T large, Ω_T^C is rejected with probability approaching unity. This is because the empirical size of the joint hypothesis test based on the sequential strategy is $\widehat{\alpha}_T = 1 - \prod_{i=1}^{T}(1 - \widehat{\alpha}_i)$, where $\widehat{\alpha}_i$ is the empirical size of the ith individual fixed time span test. As a result,

$$\lim_{T \to \infty} \widehat{\alpha}_T = \lim_{T \to \infty} 1 - \prod_{i=1}^{T}(1 - \widehat{\alpha}_i)$$
$$= 1 - \lim_{T \to \infty} \prod_{i=1}^{T}(1 - \widehat{\alpha}_i)$$
$$= 1.$$

In our Monte Carlo simulations, we illustrate this issue under a set of experiments conducted with

an increasing time span. One common approach to this problem is based on controlling the overall Family-Wise Error Rate (FWER), which ensures that no single hypothesis is rejected at a level larger than a fixed value, say α. This is typically accomplished by sorting individual p-values, and using a rejection rule which depends on the overall number of hypotheses. For further discussion, see Holm (1979), who develops modified Bonferroni bounds, White (2000), who develops the so-called "reality check", and Romano and Wolf (2005), who provide a refinement of the reality check. However, when the number of hypotheses in the composite grows with the sample size, the null will (almost) never be rejected. In other words, approaches based on the FWER are quite conservative.

An alternative approach, which allows for the number of hypotheses in the composite to grow to infinity, is based on the Expected False Discovery Rate (E-FDR). When using this approach, one controls the expected number of false discoveries (rejections). For further discussion, see Benjamini and Hochberg (1995) and Storey (2003). Although the E-FDR approach applies to the case of a growing number of hypotheses, it is very hard to implement in the presence of generic dependence across p-values, as in our context.

The above discussion, when coupled with issues of identification and test consistency, provides ample impetus for using long time span jump tests of the variety discussed in the sequel. Still, it should be noted that researchers have shown good performance of fixed time span tests over a day or a week, and we provide further evidence on this front in our Monte Carlo experiments. However, almost no one considers performing the tests over a year or even a decade. The only exception that we are aware of is Huang and Tauchen (2005). They propose using "full-sample statistics" based on BNS test statistics. They show that when the time span is long, the BNS test over-rejects the null of no realized jumps, since the approximation error on a short interval accumulates as the time span increases. Consequently, the empirical size is biased upwards.

4 Long Time Span Jump Intensity Test

Assume the existence of a sample of n observations over an increasing time span, T, and a shrinking discrete interval Δ, so that $n = \frac{T}{\Delta}$, with $T \to \infty$ and $\Delta \to 0$. Our interest lies in the following hypotheses:

$$H_0 : \lambda = 0$$

versus

$$H_A = H_A^{(1)} \cup H_A^{(2)} : \left(\lambda > 0 \text{ and } \mathrm{E}\left((Z_t - \mathrm{E}\,(Z_t))^3 \right) \neq 0 \right)$$
$$\cup \left(\lambda > 0 \text{ and } \mathrm{E}\left((Z_t - \mathrm{E}\,(Z_t))^3 \right) = 0 \right).$$

Notice that the alternative hypothesis is the union of two different alternatives, designed to allow for both symmetric and asymmetric jump size density. This setup characterizes the CSS jump test of Corradi et al. (2018) and the $CSS1$ test of Corradi et al. (2014). Moreover, both tests are based on statistics that are functions of realized third moments, or tricity.[8] However, we discuss only the $CSS1$ in the sequel, as the working paper in which that test appears will not be published, while Corradi et al. (2018) has been published.

discussed in Corradi et al. (2018). More, although their test differs in a key respect. Namely, their test is dependent on jump thresholding.

Let $Y_{k\Delta} = \ln X_{k\Delta} - \frac{\Delta}{T} \sum_{k=2}^{n} \ln X_{k\Delta}$, and $Y_{(k-1)\Delta} = \ln X_{(k-1)\Delta} - \frac{\Delta}{T} \sum_{k=2}^{n} \ln X_{(k-1)\Delta}$. Also, let

$$\widehat{\lambda}_{T,\Delta} = \frac{1}{T} \sum_{k=2}^{n} \left(Y_{k\Delta} - Y_{(k-1)\Delta} \right)^3.$$

Here, $\widehat{\lambda}_{T,\Delta}$ is the demeaned sample third moment. Consider the statistic

$$CSS1_{T,\Delta} = \frac{\sqrt{T}}{\Delta} \widehat{\lambda}_{T,\Delta}. \tag{10}$$

The asymptotic behavior of $CSS1_{T,\Delta}$ can be analyzed under the following assumption.

Assumption A: (i) $\ln X_t$ is generated by equation (1) and V_t is defined in equations (2), (3), or (4). (ii) $\ln X_t$ is generated by equation (1) and V_t is defined in equation (5). For C a generic constant, (iii) $E\left(|V_t|^k\right) \leq C$, $k \geq 3$, (iv)N_t satisfies equations (6)-(8), and λ_t is either constant, or satisfies equation (9). (v) The jump size, Z_t, is independently and identically distributed, and $E\left(|Z_t|^k\right) \leq C$, for $k \geq 6$.

Corradi et al. (2014) show that under assumptions A(i) and A(iii)-(v), assuming that as $n \to \infty$, $T \to \infty$ and $\Delta \to 0$, then under H_0 : $CSS1_{T,\Delta} \xrightarrow{d} N(0, \omega_0)$, with $\omega_0 = 15E\left(V_{k\Delta}^3\right) + 4\left(\mathrm{E}\left(V_{k,\Delta}\right)\right)^3 - 12E\left(V_{k,\Delta}\right)E\left(V_{k,\Delta}^2\right)$. Also, they prove that under $H_A^{(1)}$, there exists an $\varepsilon > 0$, such that: $\lim_{T\to\infty,\Delta\to0} \Pr\left(\frac{\Delta}{\sqrt{T}} |CSS1_{T,\Delta}| > \varepsilon\right) = 1$; and under $H_A^{(2)}$, there exists an $\varepsilon > 0$, such that: $\lim_{T\to\infty,\Delta\to0} \Pr\left(\Delta |CSS1_{T,\Delta}| > \varepsilon\right) = 1$.

It follows immediately that $CSS1_{T,\Delta}$ converges to a normal random variable under the null hypothesis, diverges at rate $\frac{\sqrt{T}}{\Delta}$ under the alternative of asymmetric jumps, and diverges at the slower rate of $\frac{1}{\Delta}$ under the alternative of symmetric jumps.

Given that the variance is of a different order of magnitude under the null and under each alternative, the "standard" nonparametric bootstrap is not asymptotically valid. This issue arises because the variance of the bootstrap statistic mimics the sample variance. This implies that the

---

[8]The key difference between the CSS and $CSS1$ tests is that the former utilized thresholding, while the latter does not.

bootstrap statistic is of order Δ^{-1} under the alternative. This is not be a problem under $H_A^{(1)}$, since the statistic is of order $\sqrt{T}\Delta^{-1}$, but is a problem under $H_A^{(2)}$, since the actual and bootstrap statistics would be of the same order. To ensure power against $H_A^{(2)}$, it suffices to ensure that the bootstrap statistic is of a smaller order than the actual statistic. This can be accomplished by resampling observations over a rougher grid, $\widetilde{\Delta}$, using the same time span, T.

Set the new discrete interval to be $\widetilde{\Delta}$, such that $\Delta/\widetilde{\Delta} \to 0$, and resample, with replacement, $\left(Y_{k\widetilde{\Delta}}^* - Y_{(k-1)\widetilde{\Delta}}^*, ..., Y_{\widetilde{n}\widetilde{\Delta}}^* - Y_{(\widetilde{n}-1)\widetilde{\Delta}}^*\right)$ from $\left(Y_{k\widetilde{\Delta}} - Y_{(k-1)\widetilde{\Delta}}, ..., Y_{\widetilde{n}\widetilde{\Delta}} - Y_{(\widetilde{n}-1)\widetilde{\Delta}}\right)$, where $\widetilde{n} = T/\widetilde{\Delta}$. Now, let

$$\widetilde{\lambda}_{T,\widetilde{\Delta}} = \frac{1}{T}\sum_{k=2}^{\widetilde{n}}\left(Y_{k\widetilde{\Delta}} - Y_{(k-1)\widetilde{\Delta}}\right)^3,$$

and

$$\widetilde{\lambda}_{T,\widetilde{\Delta}}^* = \frac{1}{T}\sum_{k=2}^{\widetilde{n}}\left(Y_{k\widetilde{\Delta}}^* - Y_{(k-1)\widetilde{\Delta}}^*\right)^3.$$

Further, define the bootstrap statistic as

$$CSS1_{T,\widetilde{\Delta}}^* = \frac{\sqrt{T}}{\widetilde{\Delta}}\left(\widetilde{\lambda}_{T,\widetilde{\Delta}}^* - \widetilde{\lambda}_{T,\widetilde{\Delta}}\right).$$

Finally, let $c_{\alpha,B,\Delta,\widetilde{\Delta}}^*$ and $c_{(1-\alpha),B,\Delta,\widetilde{\Delta}}^*$ be the $(\alpha/2)^{th}$ and $(1-\alpha/2)^{th}$ critical values of the empirical distribution of $CSS1_{T,\widetilde{\Delta}}^*$, constructed using B bootstrap replications. Corradi et al. (2014) show that under assumptions A(i) and A(iii)-(v), and assuming that as $n \to \infty$, $B \to \infty$, $T \to \infty$, $\Delta \to 0$, $\widetilde{\Delta} \to 0$ and $\Delta/\widetilde{\Delta} \to 0$, then under H_0 :

$$\lim_{T,B\to\infty,\Delta,\widetilde{\Delta}\to 0}\Pr\left(c_{\alpha/2,B,\Delta,\widetilde{\Delta}}^* \leq CSS1_{T,\Delta} \leq c_{(1-\alpha/2),B,\Delta,\widetilde{\Delta}}^*\right) = 1-\alpha;$$

and under $H_A^{(1)} \cup H_A^{(2)}$:

$$\lim_{T,B\to\infty,\Delta,\widetilde{\Delta}\to 0}\Pr\left(c_{\alpha/2,B,\Delta,\widetilde{\Delta}}^* \leq CSS1_{T,\Delta} \leq c_{(1-\alpha/2),B,\Delta,\widetilde{\Delta}}^*\right) = 0.$$

It is immediate to see that rejecting the null whenever $\frac{\sqrt{T}}{\Delta}\widehat{\lambda}_{T,\Delta} < c_{\alpha/2,B,\Delta,\widetilde{\Delta}}^*$ or $\frac{\sqrt{T}}{\Delta}\widehat{\lambda}_{T,\Delta} > c_{(1-\alpha/2),B,\Delta,\widetilde{\Delta}}^*$, and otherwise failing to reject, delivers a test with asymptotic size equal to α and asymptotic power equal to unity. Note that the bootstrap statistic is of P^*-probability order $\frac{1}{\widetilde{\Delta}}$ under both $H_A^{(1)}$ and $H_A^{(2)}$, while the actual statistic is of $P-$probability order $\frac{\sqrt{T}}{\Delta}$ under $H_A^{(1)}$ and $\frac{1}{\Delta}$ under $H_A^{(2)}$. Hence, the condition that $\Delta/\widetilde{\Delta} \to 0$ ensures unit asymptotic power under $H_A^{(2)}$.

The *CSS1* test is not robust to leverage. In particular, the results presented above rely on the fact that under the null of no jumps, returns are symmetrically distributed. More precisely, all results are derived under the assumption that $\mathrm{E}\left(\left(Y_{k\Delta} - Y_{(k-1)\Delta}\right)^3\right) = 0$, whenever

there are no jumps. However, in the presence of leverage, (i.e. V_t is generated as in equation (5)), $\mathrm{E}\left(\left(\int_{(k-1)\Delta}^{k\Delta} V_s^{1/2} \mathrm{d}W_{1,s}\right)^3\right) \neq 0$, and is instead of order Δ^2. For example, if V_t is generated by a square root process (i.e., $\mathrm{d}V_t = \kappa\left(\theta - V_t\right)\mathrm{d}t + \eta V_t^{1/2}\mathrm{d}W_{2,t}$), then $\mathrm{E}\left(\left(Y_{k\Delta} - Y_{(k-1)\Delta}\right)^3\right) = \lambda\mathrm{E}\left(Z_t - \mathrm{E}\left(Z_t\right)\right)^3 \Delta + \frac{\eta\theta\rho}{2\kappa}\Delta^2$ (see Aït-Sahalia et al. (2015)). Although, the contribution to the third moment of the asymmetric jump component is of a larger order than that of the leverage component, inference based on the comparison of $CSS1_{T,\Delta}$ with the bootstrap critical values $c^*_{\alpha,B,\Delta,\widetilde{\Delta}}$ and $c^*_{(1-\alpha),B,\Delta,\widetilde{\Delta}}$ will lead to the rejection of the null of no jumps, even if the null is true. To avoid spurious rejection due to the presence of leverage, use the following modified statistic:

$$\widetilde{CSS1}_{T,\Delta} = \frac{1}{T^{1/2+\varepsilon}} CSS1_{T,\Delta}, \tag{11}$$

with $\varepsilon > 0$, arbitrarily small. For this test statistic, Corradi et al. (2014) show that under assumptions A(ii)-(v) hold, and assuming that as $n \to \infty$, $T \to \infty$, $\Delta \to 0$, $\widetilde{\Delta} \to 0$, and $(T^{1/2+\varepsilon}\Delta)/\widetilde{\Delta} \to 0$, then under H_0 :

$$\lim_{T,B\to\infty,\Delta,\widetilde{\Delta}\to 0} \Pr\left(c^*_{\alpha/2,B,\Delta,\widetilde{\Delta}} \leq \widetilde{CSS1}_{T,\Delta} \leq c^*_{(1-\alpha/2),B,\Delta,\widetilde{\Delta}}\right) = 1; \tag{12}$$

and under $H_A^{(1)} \cup H_A^{(2)}$:

$$\lim_{T,B\to\infty,\Delta,\widetilde{\Delta}\to 0} \Pr\left(c^*_{\alpha/2,B,\Delta,\widetilde{\Delta}} \leq \widetilde{CSS1}_{T,\Delta} \leq c^*_{(1-\alpha/2),B,\Delta,\widetilde{\Delta}}\right) = 0.$$

It follows that inference based on the comparison of $\widetilde{CSS1}_{T,\Delta}$ with the bootstrap critical values $c^*_{\alpha,B,\Delta,\widetilde{\Delta}}$ and $c^*_{(1-\alpha),B,\Delta,\widetilde{\Delta}}$ delivers a test with zero asymptotic size and unit asymptotic power.

There are two major differences between the CSS test developed in Corradi et al. (2018) and the $CSS1$ type tests outlined above. First, instead of using an adjustment term in equation (11) to account for the leverage effect, Corradi et al. (2018) use a longer time span, T^+, with $\kappa = T^+/T \to \infty$, to eliminate the leverage effect. Specifically, now assume the existence of a sample of n^+ observations over an increasing time span, T^+, and a shrinking discrete interval Δ, so that $n^+ = \frac{T^+}{\Delta}$, with $T^+ \to \infty$ and $\Delta \to 0$. Define $n = T/\Delta = n^+ - \frac{T^+ - T}{\Delta}$, with $T^+/T \to \infty$. Let

$$\widehat{\mu}_{T,\Delta} = \frac{1}{T} \sum_{k=1}^{n-1} \left(\ln X_{(k+1)\Delta} - \ln X_{k\Delta} - \frac{\ln X_{n\Delta} - \ln X_\Delta}{n}\right)$$

$$-\frac{1}{T^+} \sum_{k=1}^{n^+-1} \left(\ln X_{(k+1)\Delta} - \ln X_{k\Delta} - \frac{\ln X_{n\Delta} - \ln X_\Delta}{n^+}\right)\mathbf{1}_{\{|\ln X_{(k+1)\Delta} - \ln X_{k\Delta}| \leq \alpha\Delta^\varpi\}}$$

where the second term on the right-hand side correct for the leverage effect.

11

Define the statistic,

$$\widehat{CSS}_{T,\Delta} = \frac{sqrt(T)}{\Delta}\widehat{\mu}_{T,\Delta}.$$

The $\widehat{CSS}_{T,\Delta}$ is evidently closely related to the $CSS1_{T,\Delta}$ test in equation (10).

Second, instead of using a rescaled bootstrap to account for the fact that the variance of the statistic is of a different order of magnitude under the null and under each alternative, Corradi et al. (2018) introduce a threshold variance estimator to consistently estimate the variance of $\widehat{CSS}_{T,\Delta}$ under the null and bounded in probability under the union of alternatives. Specifically, define,

$$\widehat{\sigma}^2_{T,\Delta} = \frac{1}{T\Delta^2}\sum_{k=1}^{n-1}\left(\ln X_{(k+1)\Delta} - \ln X_{k\Delta} - \frac{\ln X_{n\Delta} - \ln X_\Delta}{n}\right)^3 \mathbf{1}_{\{|\ln X_{(k+1)\Delta} - \ln X_{k\Delta}| \le \alpha\Delta^\varpi\}}$$

then the following t-statistic can be used for inference,

$$\widehat{t}_{T,\Delta} = \frac{\widehat{CSS}_{T,\Delta}}{\widehat{\sigma}_{T,\Delta}}$$

Corradi et al. (2018) show that under certain assumptions, $\widehat{t}_{T,\Delta}$ converges to a standard normal variable under the null, diverges at rate $\frac{\sqrt{T}}{\Delta}$ under $H_A^{(1)}$, and diverges at the slower rate of $\frac{1}{\Delta}$ under $H_A^{(2)}$.

5 Fixed Time Span Realized Jump Tests

In this section, we briefly review three fixed time span realized jump tests that are evaluated in our Monte Carlo and empirical experiments. These tests are the *ASJ*, *BNS* and *PZ* tests discussed above.

5.1 Aït-Sahalia and Jacod (*ASJ*: 2009) Test

Aït-Sahalia and Jacod (2009) propose a jump test based on calculating the ratio between two realized higher order power variations with different sampling intervals Δ and $q\Delta$, respectively, where q is an integer chosen prior to test construction. The pth order realized higher order power variation is defined as follows,

$$\widehat{B}(p,\Delta) = \sum_{k=2}^{\lfloor 1/\Delta \rfloor} |\ln X_{k\Delta} - \ln X_{(k-1)\Delta}|^p \tag{13}$$

where $\lfloor \chi \rfloor$ indicates the integer part of χ.

The ratio between two realized power variations with different sampling intervals is then,

$$\widehat{S}(p,q,\Delta) = \frac{\widehat{B}(p,q\Delta)}{\widehat{B}(p,\Delta)}. \tag{14}$$

The test statistic is defined as,

$$ASJ = \frac{q^{\frac{p}{2}-1} - \widehat{S}(p,q,\Delta)}{\sqrt{V_n^c}}, \tag{15}$$

where in the denominator, V_n^c, can be estimated using a truncated estimator,

$$\widehat{V}_n^c = \Delta \frac{\widehat{A}(2p,\Delta)M(p,q)}{\widehat{A}(p,\Delta)^2}, \tag{16}$$

where

$$M(p,q) = \frac{1}{\mu_p^2}(q^{p-2}(1+q)\mu_{2p} + q^{p-2}(q-1)\mu_p^2 - 2q^{p/2-1}\mu_{q,p}),$$

and $\widehat{A}(p,\Delta)$ is defined as follows,

$$\widehat{A}(p,\Delta) = \frac{\Delta^{1-p/2}}{\mu_p} \sum_{k=2}^{\lfloor 1/\Delta \rfloor} |\ln X_{k\Delta} - \ln X_{(k-1)\Delta}|^p \mathbf{1}_{\{|\ln X_{k\Delta} - \ln X_{(k-1)\Delta}| \leq \alpha \Delta^{\varpi}\}}, \tag{17}$$

where $\alpha > 0$, $\varpi \in (0,1/2)$, and $\alpha\Delta^{\varpi}$ functions as a threshold separating the continuous component from the jump component. Alternatively, V_n^c, can be estimated using a multipower variation estimator,

$$\widetilde{V}_n^c = \Delta \frac{M(p,q)\widetilde{A}(p/(\lfloor p \rfloor + 1), 2\lfloor p \rfloor + 2, \Delta)}{\widetilde{A}(p/(\lfloor p \rfloor + 1), \lfloor p \rfloor + 1, \Delta)^2}, \tag{18}$$

where

$$\widetilde{A}(r,l,\Delta) = \frac{\Delta^{1-lr/2}}{\mu_r^l} \sum_{k=l}^{\lfloor 1/\Delta \rfloor - l + 1} \prod_{j=0}^{l-1} |\ln X_{(k+j)\Delta} - \ln X_{(k+j-1)\Delta}|^r, \tag{19}$$

and

$$\mu_r = \mathbb{E}(|U|^r) \quad \text{and} \quad \mu_{q,p} = \mathbb{E}(|U|^p|U + \sqrt{q-1}V|^p),$$

for $U, V \overset{i.i.d}{\sim} N(0,1)$.

In practice, for any significance level α, if $ASJ > Z_{\alpha}$, where Z_{α} is the $(1-\alpha)^{th}$ quantile of the standard normal distribution, one rejects the null of no jumps on the fixed interval $[0, 1]$.

5.2 Barndorff-Nielsen and Shephard (BNS: 2006) Test

The Barndorff-Nielsen and Shephard (2006) test compares the difference between two estimators of integrated volatility: one which is robust to jumps and the other which is not, to test for jumps on

a particular sample path. Barndorff-Nielsen and Shephard (2004) introduce the realized bipower variation (BPV) which is a robust estimator of the integrated volatility. Namely, they consider

$$BPV = \frac{\pi}{2}(\frac{m}{m-1}) \sum_{k=2}^{\lfloor 1/\Delta \rfloor} |\ln X_{(k+1)\Delta} - \ln X_{k\Delta}||\ln X_{k\Delta} - \ln X_{(k-1)\Delta}|, \tag{20}$$

where $m = \lfloor 1/\Delta \rfloor$. Realized BPV is a special case of the following realized multipower variation with $p = 2$,

$$MPV(p) = \mu_{\frac{2}{p}}^{-p}(\frac{m}{m-p+1}) \sum_{k=p}^{\lfloor 1/\Delta \rfloor - p+1} \prod_{j=0}^{p-1} |\ln X_{(k+j)\Delta} - \ln X_{(k+j-1)\Delta}|^{\frac{2}{p}}. \tag{21}$$

In this paper, we analyze the following version of their test statistic:

$$BNS = \Delta^{-\frac{1}{2}} \frac{1 - \frac{BPV}{RV}}{\sqrt{((\frac{\pi}{2})^2 + \pi - 5)\max(1, \frac{TPV}{(BPV)^2})}}, \tag{22}$$

where RV is the realized volatility (i.e., the sum of squared high-frequency returns), and TPV is tripower variation (i.e., $MPV(3)$).

The authors prove that under the null, $BNS \xrightarrow{d} N(0,1)$. As a result, one rejects the null of no jumps on some fixed interval $[0, 1]$, if the test statistic $BNS > Z_\alpha$.

5.3 Podolskij and Ziggel (PZ: 2010) Test

Podolskij and Ziggel (2010) modify the original truncated power variation statistic proposed in Mancini (2009) by introducing a sequence of positive $i.i.d.$ random variables $\{\eta_i\}_{i \in [1, \lfloor 1/\Delta \rfloor]}$, with expectation one and finite variance. Namely, they consider

$$T(\ln X, p) = \Delta^{\frac{1-p}{2}} \sum_{k=2}^{\lfloor 1/\Delta \rfloor} |\ln X_{k\Delta} - \ln X_{(k-1)\Delta}|^p (1 - \eta_i \mathbf{1}_{\{|\ln X_{k\Delta} - \ln X_{(k-1)\Delta}| \le \alpha \Delta^\varpi\}}). \tag{23}$$

The test statistic that they propose has the following form,

$$PZ = \frac{T(\ln X, p)}{Var^*(\eta)\widehat{A}(2p, \Delta)}, \tag{24}$$

where $\widehat{A}(2p, \Delta)$ is the original truncated power variation in equation (17). The authors prove that under the null of no jumps, $PZ \xrightarrow{d} N(0,1)$, and explodes under the alternative. As a result, one rejects the null if $PZ > Z_\alpha$.

14

6 Monte Carlo Simulations

In this section, we report the results of Monte Carlo experiments used to analyze the finite sample properties of the tests introduced above. The simulations are designed to show: (i) the relevance of inconsistency of the fixed time span tests, when tested against non-zero jump intensity in the underlying DGP; (ii) the relevance (or lack thereof) of sequential testing bias when performing daily jump tests, sequentially, along sample paths with a long time span; (iii) the empirical size and power of fixed time span jump tests when applied directly to samples with long time spans; and (iv) the finite sample properties of the $CSS1$, $\widetilde{CSS1}$ and CSS tests.

The DGP under the null hypothesis in all simulations is the following stochastic volatility model,

$$
\ln X_t = \ln X_0 + \int_0^t \bar{\mu} ds + \int_0^t \sigma_s dW_s,
$$
$$
\sigma_t^2 = \sigma_0^2 + \kappa_\sigma \int_0^t (\bar{\sigma}^2 - \sigma_s^2) ds + \zeta \int_0^t \sqrt{\sigma_s^2} dB_s,
$$

(25)

where the stochastic volatility follows a square root process.[9] Leverage effects are characterized by $\text{corr}(dW_s, dB_s) = \rho$, where $\rho = \{0, -0.5\}$. Under the alternative, we simulate jumps as a compound Poisson process. Namely, we add $\sum_{i=1}^{N_t} J_i$ to the null DGP, where N_t is a Poisson process characterized by intensity parameter λ, which determines the frequency of jump arrivals, and J_i is independently and identically drawn from a normal distribution, which characterizes the jump size. All parameter values for various DGP permutations considered are given in Table 1. Of note is that the parameter values used in our experiments are chosen to regions of the parameter space where the tests shift from having strong finite sample properties to having weaker finite sample properties. Thus, for example, while we broadly mimic the parameterizations used in the extant literature (see e.g., Huang and Tauchen (2005) and Aït-Sahalia and Jacod (2009), in some cases, our parameters are slightly smaller. For example, Huang and Tauchen (2005) have jump magnitude standard deviation parameters ranging from 0.5 to 2.0, while ours range from 0.25 to 1.25. The sampling frequency in our simulations is 5-minute (i.e., 78 observations per day).[10] Using the Milstein discretization scheme, we simulate log-price sample paths over $T = 500$ days, so that there are 39000 observations in total, for each sample path. Simulation results are calculated based on 1000 replications, and tests are implemented using 0.05 and 0.10 significance levels.

Table 2 reports empirical size of daily fixed time span jump tests. In this table, however, the test is applied in two different ways. For entries under the "Jump Days" header, the empirical

---

[9]Note that in all of our experiments, we do not consider jumps in the volatility process, as mentioned in Section 2.

[10]In all experiments, we first simulate asset log-prices using a Milstein approximation scheme, with a finer interval $h = 1/312$. We then construct 5-minute returns, i.e. with $\Delta = 1/78$. Other numerical simulation strategies can also be used. For example, one can first simulate 1-second frequency data, then sample the data at a 5-minute frequency, as in Christensen et al. (2014). However, considering the great computational burden associated with simulating long-span data, this is left to future research.

size of the daily tests are reported. One can think of these experiments as reporting rejection frequencies of 500,000 tests (since $T = 500$ and there are 1000 Monte Carlo replications). For entries under the "Sequential Testing Bias" header, sequences of T tests (corresponding the the length of our daily samples) are run, and overall rejection frequencies across all T tests are reported, where T ranges from 1 to 500 days. Thus, these entries indicate the accumulation of sequential testing bias associated with repeated application of the tests across multiple days. Turning first to the "Jump Days" empirical size results, it is evident that the BNS test is least favorably sized, as expected, while the ASJ and PZ tests are very accurately sized, across all DGPs, with the only exception being that the PZ test is undersized when $T = 5$. Now, consider the "Sequential Testing Bias" results in the table. As expected, sequential testing bias leads to a 1.000 rejection rate as T increases beyond 50 days, and these rejections rates are achieved surprisingly quickly, as T increases, although it is interesting to note that the PZ test suffers from slightly less bias, for smaller values of T.[11]

Table 3 reports empirical power of daily fixed time span jump tests, defined as the rejection rate of daily jump tests across each individual day in each sample path, averaged across all 1000 replications. As in Table 2, one can think of these experiments as reporting rejection frequencies of 500,000 tests. Interestingly, power is often small, even when $\lambda = 0.4$, which is a relatively large value, for finite-activity jumps. Among the three jump tests, the ASJ test has the lowest power, while BNS and PZ test are somewhat better. In interpreting these results, note that, intuitively speaking, the empirical power of daily jump tests against non-zero jump intensity is largely determined by the magnitude of the jump intensity, since this parameter determines the frequency or probability of jump arrivals. Even if these tests have good power against jumps when they occur, for daily intervals without any jumps, it is not surprising to observe that these tests do not reject the null in favor of non-zero jump intensity. Therefore, as long as the intensity is finite, the probability of jumps not occurring on a particular fixed interval is positive, which in turn affects the empirical power of all fixed time span jump tests. However, the tests are also clearly impacted by jump size magnitude. For example, when σ increase from 0.25 to 1.25 (compare $DGPs$ 3 and 4 with $DGPs$ 5 and 6 – symmetric jumps, or compare $DGPs$ 7 and 8 with $DGPs$ 9 and 10 – asymmetric jumps), in which cases, empirical power increases by around 30% under symmetric jumps. The exception is the BNS and PZ tests, which show little power improvement, under the asymmetric jump case. However, there is still a trade-off between the three tests, as the ASJ test has overall less power for the case of symmetric jumps.

Tables 4–7 report findings from experiments where the "entire" sample of T days was used in a single application of the fixed time span tests. This testing strategy is of interest, because there is no reason that fixed time span tests need be implemented using only one day worth of data;

---

[11]It would be interesting to analyze results for the case where fixed time span tests are sequentially conducted, but the FWER is controlled for, as discussed in Section 3. This is left to future research.

and when they are implemented in this manner, they constitute a direct alternative to the use of our long time span tests. First, turn to Table 4, where empirical size is reported. Among the three tests, ASJ is the clear winner, with size remaining stable even when $T = 500$. This is an interesting and surprising result, suggesting the broad usefulness of the ASJ test. The BNS and PZ tests perform as expected, on the other hand. For example, the empirical size of the PZ test approaches unity very quickly, and is already approximately 0.5, even for $T = 50$, indicating the weak ability of this test to control size as the time span increases. As expected, and as can be seen upon inspection of Table 5–7, the empirical power of all three tests approaches unity quickly as T increases. For example, the empirical power of the ASJ test are over 0.9 for almost all $DGPs$, when $T = 50$. In summary, the ASJ test is well sized and has great power under long time span testing. This test, thus, is a clear alternative to the long time span tests discussed in the sequel.

Tables 8–12 contain the results of our experiments run using long time span jump tests (i.e. the $CSS1$, $\widetilde{CSS1}$ and CSS tests). As discussed above, the leverage-robust $CSS1$ test (i.e., $\widetilde{CSS1}$) sacrifices power in order to ensure robustness against leverage effects. Specifically, in equation (11), due to the extra term, $\frac{1}{T^{1/2+\epsilon}}$, the leverage-robust test statistic diverges at rate $\frac{1}{T^\epsilon\Delta}$ and $\frac{1}{T^{1/2+\epsilon}\Delta}$, under the alternatives of asymmetric and symmetric jumps, respectively. In practice, the sampling interval, Δ, is usually small and fixed, and the test statistic under the alternatives shrinks with an increasing time span, particularly when jumps are symmetric, while the bootstrapped critical values are of order $1/\widetilde{\Delta}$. As a result, when constructing the leverage-robust test, we propose a rule-of-thumb called our "T-varying" strategy, in order to choose the subsampling interval, $\widetilde{\Delta}$, used in bootstrapping (see notes to Table 8 for details). This rule-of-thumb results in improved power in our experiments. However, it is an *ad hoc* data driven method, and further research into its properties remains to be done. Summarizing, we utilize coarser $\widetilde{\Delta}$, as T grows. Some choices of $\widetilde{\Delta}$ in our experiments can be sub-optimal. For instance, consider $\widetilde{\Delta} = 2$ and 3.2 when $T = 300$ and 500, respectively. However, these choices appear sub-optimal only when T is very long. Thus, the subsamples used are still adequate for bootstrap resampling. Since a coarser value for $\widetilde{\Delta}$ (i.e., a larger subsampling interval) diminishes the magnitude of our bootstrap critical values, this strategy reduces the magnitude of decreases in power that are due to the adjustment term being inversely proportional to T. The effect of this *ad hoc* data driven method for improving power on empirical size is found to be negligible, and hence the method is utilized in all of our leverage-robust testing experiments, and later in our empirical analysis.

Turning to the results reported in these tables, first consider the empirical size of the $CSS1$ and $\widetilde{CSS1}$ tests (see Table 8). It is immediately apparent that the $CSS1$ test has good empirical size for $DGP~1$ (i.e., the "no leverage" case). However, and as expected, size diverges when there is leverage. Again as expected (see equation (12)), $\widetilde{CSS1}$ has zero empirical size, regardless of the presence of leverage, for values of T greater than 5. Interestingly, though when $T = 5$, the test is approximately correctly sized; thus indicating that our long time span test is an alternative to the

short time span tests discussed earlier for small values of T. Of course, T should clearly not be equal to one for the application of the long time span tests. In Table 9, we report the empirical size of the CSS test, for different permutations of T and Δ. We find several interesting results. First, rejection frequencies are lower when $\kappa = T^+/T$ is closer to unity, especially for shorter T, given Δ.[12] This is not surprising because for $T^+ = T$ the test statistic is degenerate under the null (see Corradi et al. (2018)). Thus, it is advisable to use a reasonably large value for κ, such as $\kappa = 10$, given the assumption that $T^+/T \to \infty$. Next, consider the results for the case where $\rho = 0$. The CSS test is correctly sized for $T = 60, 70, 80$, when $\Delta = 1/78$, and is slightly undersized even for $T = 220$, when $\Delta = 1/156$. This finding suggests that the ratio of T to Δ is extremely crucial to the finite sample performance of the CSS test. Again, this is not surprising given the key assumption that $1/\Delta < T < 1/\Delta^2$. Finally, we observe that the test becomes oversized even faster when there is leverage (i.e. the case with $\rho = -0.5$). However, as long as T and Δ are carefully chosen, the finite sample performance of the CSS test is adequate.

Next, notice in Table 10 that the empirical power of the $CSS1$ is good across all cases. This result even holds for the case where jumps are symmetric and $\sigma = 0.25$ (i.e., $DGPs$ 3 and 4), where power is still reasonable, although tests are naturally less powerful than the case where jumps are asymmetric. Turning to Table 11, where empirical power of the $\widetilde{CSS1}$ is reported. As expected, empirical power is sacrificed, particularly when jumps are symmetric and $\sigma = 0.25$ (i.e., $DGPs$ 3 and 4). However, when $\sigma = 1.25$ (i.e., $DGPs$ 5 and 6), this sacrifice is substantially reduced, and power is reasonably good in almost all cases, even when λ is small. Finally, the empirical power of the CSS test is reported in Table 12.[13] The CSS test is much more powerful than the $\widetilde{CSS1}$ test for all cases. For example, when jumps are less frequent and are symmetric and very weak (i.e. $DGPs$ 3 and 4 with $\lambda = 0.1$), the power of CSS test is over 47%, when $T = 50$ and $\Delta = 1/78$, at a 10% nominal level, while the power of the $\widetilde{CSS1}$ test is less than 20%. Furthermore, the power of the CSS test increases as T grows, for a fixed Δ. In contrast, as discussed above, the power of the $\widetilde{CSS1}$ test decreases as T grows, for fixed Δ. Coupled with our earlier findings concerning size, we thus have strong evidence that the $\widetilde{CSS1}$ and the CSS tests are adequate tests for evaluating the presence of jumps in long time spans.

---

[12]We assume that the sample over an increasing time span, T^+ used in the construction of the test, is observed and we examine a subsample with time span, T. For more details, see Corradi et al. (2018).

[13]We only report results associated with less frequent and weak jumps (i.e. $\lambda = 0.1$ and 0.4) and for jump sizes that are $i.i.d$ normally distributed with $\mu = 0$, $\sigma_J = 0.25$ and 1.25 (i.e., small and symmetric jumps, corresponding to the "worst" alternatives considered). Complete results are available upon request.

7   Empirical Examination of Stock Market Data

7.1   Data

We analyze intraday TAQ stock price data sampled at a 5-minute frequency, for the period including observations from the beginning of 2006 through 2013.[14] In particular, we examine: (i) twelve individual stocks including American Express Company (AXP), Bank of America Corporation (BAC), Cisco Systems, Inc. (CSCO), Citigroup Inc. (C), The Coca-Cola Company (KO), Intel Corporation (INTC), JPMorgan Chase & Co. (JPM), Merck & Co., Inc. (MRK), Microsoft Corporation (MSFT), The Procter & Gamble Company (PG), Pfizer Inc. (PFE) and Wal-Mart Stores, Inc. (WMT)); (ii) nine sector ETFs including Materials Select Sector SPDR ETF (XLB), Energy Select Sector SPDR ETF (XLE), Financial Select Sector SPDR ETF (XLF), Industrial Select Sector SPDR ETF (XLI), Technology Select Sector SPDR ETF (XLK), Consumer Staples Select Sector SPDR ETF (XLP), Utilities Select Sector SPDR ETF (XLU), Health Care Select Sector SPDR ETF (XLV) and Consumer Discretionary Select Sector SPDR ETF (XLY); and (iii) the SPDR S&P 500 ETF (SPY). Overnight returns are excluded from our dataset.

7.2   Empirical Findings

Turning our discussion first to Figures 1–2, note that the bar charts in these figures depict annual ratios of jump days for all of our stocks and ETFs, based on application of the *ASJ*, *BNS*, and *PZ* tests (see legend to Figure 1). For example, 0.2 indicates that there were jumps founds on 20% of the trading days in a given year. As expected, jumps are widely detected in asset prices and indexes over almost any year. Sometimes, the annual percentage of jump days even appears to be inconceivably large, at near 50%. Additionally, while the alternative tests often perform similarly (e.g. all three testing methods find jumps during around 40% of the days in 2006 for XLU and XLP), there are substantial differences for some stocks (e.g. in 2013 the *PZ* tests detects jumps twice as frequently as the other fixed time span tests).

As expected, the *ASJ* test is the most conservative among the three tests. In almost all cases, the *ASJ* test detects the fewest number of "jump days". For instance, in 2008, *ASJ* test only finds 7.5% jump days for XLK, while the *PZ* and *BNS* tests find jumps on 17.4% and 22.5% of days, respectively. For SPY, the *ASJ* test finds around 1/3 as many jumps as the other tests, in 2009. This finding is consistent with evidence from our Monte Carlo experiments (see Table 3). However, even with the most conservative test, we regularly detect over 15% jump days for many assets, including XLV for 2006 through 2010, XLB and XLY in 2006, and XLF and XLI for 2006 and 2007. Additionally, jump-day percentages are generally larger for our sector ETFs than for individual

---

[14]It will be interesting to use data sampled at a higher frequency, as suggested in Christensen et al. (2014), and to carry out inference on the relative discovery rate of the jump tests, as recommended by a referee. However, given that the tests considered in this paper do not address microstructure noise, this is left to future research.

stocks. This is not surprising, considering that the individual stocks that we consider are all much larger than the sector ETFs, in terms of the trading volume, for the periods from 2006 to 2013. Still, it is also apparent, upon inspection of the figures, that the percentage of jumps detected in our ETFs is declining over time, on an annual basis. This pattern does not characterize individual stocks, however. We conjecture that a possible reason for this is that sector ETFs where not as frequently traded in the early years of our sample. For instance, typical daily trading volume for XLP or XLY was around 1 million, including pre-market trading and after-hours trading volumes, between 2006 and 2008. This volume is around 1% to 10% of the trading volume of BAC, and 0.15% to 2.5% of the trading volume of SPY, over the same period.

We now turn to a discussion of the results tabulated in Tables 13–19. In these tables, jump tests results based on the examination of long time spans are reported for the *ASJ*, *CSS1*, $\widetilde{CSS1}$ and *CSS* tests. In these tables, tabulated entries are test statistics, and those entries with *, **, and *** indicate rejections of the "no-jump" null at 0.10, 0.05, and 0.01 significance levels, respectively. In these tables, the "long span" considered is one year for the *ASJ*, *CSS1* and $\widetilde{CSS1}$, corresponding to the period of time for which annual jump-day ratios were reported in Figures 1–2, and is one quarter for the *CSS* test. Consider first the results of the *ASJ* test reported in Table 13 for our ETFs. Interestingly, there are various ETFs for which no jumps are found. For example, for XLE, no jumps are found in 2006, 2008, 2010, and 2011. In 2011, no jumps are found for 7 of 10 ETFs. Still, in 2007, jumps are found for all 9 ETFs, and in 2008, jumps are found for 7 ETFs. Thus, the evidence concerning jumps appears much more nuanced when the *ASJ* tests is utilized using long time spans. Of course, however, the discussion above concerning trading volume effects during the early years of our sample still applies. Thus, it is difficult to be sure whether the increase in the frequency of jumps found in earlier years for our sector ETFs is an indicator of the ensuing financial collapse of 2008, or whether this finding is simply an artifact of the data. We leave further investigation of this issue to future research.

Turning to Tables 14 and 15, which again report on ETFs, note that these tables include results for the *CSS1* (Table 14) and $\widetilde{CSS1}$ (Table 15) tests. As expected, given our Monte Carlo findings, and assuming the presence of leverage, rejections based on the *CSS1* test are not only frequent, but are actually more frequent than rejections based on the *ASJ* test. Indeed, given the presence of leverage, these results carry little weight. However, we know that the $\widetilde{CSS1}$ performs adequately, given the presence of leverage. It is perhaps not surprising, then, that the number of years for which jumps are found decreases substantially when $\widetilde{CSS1}$ is used, relative to when testing using *CSS1*. Indeed, in Table 15, note that there are many ETFs for which no jumps are found across multiple different years. Still, it should be stressed that while $\widetilde{CSS1}$ is robust to the presence of leverage, the cost of making it thus is a reduction in power, as discussed in the previous sections of this paper. Thus, our conjecture is that the "truth" likely lies somewhere between the results reported based on application of the *ASJ* and the $\widetilde{CSS1}$ tests. Still, either way, it is clear that application

of long time span tests results in fewer findings of jumps. It is this feature of the tests that is most intriguing, given its implications on the specification and estimation of diffusion models.

Tables 16–18 contain results that are analogous to those reported in Tables 13–15, except that individual stocks are analyzed. Interestingly, the test rejection patterns that appear upon inspection of the entire in these tables confirms our above discussion based on ETF analysis. Namely, there are various years for which no jumps are found based on application of the ASJ test, and this incidence of "non-rejections" increases when one utilizes the $\widetilde{CSS1}$ test.

Finally, Table 19 includes results based on the application of the CSS test.[15] We observe more rejections in this table, which is not surprising, given the results of our Monte Carlo experiments, which indicate that the CSS test is much more powerful than the $\widetilde{CSS1}$ test, even when jumps are infrequent and weak. However, there is a possibility that some of these rejections are spurious, due to the fact that when there is leverage, the test bias associated with the CSS test increases with T, for a fixed sampling frequency, at a faster rate than does the bias associated with the $\widetilde{CSS1}$ test. Regardless, when the CSS is used, we once again find multiple quarters that exhibit no evidence of jumps.

In summary, we conclude that the usual "toolbox" used by financial econometricians might be usefully augmented by including in it long time span jump tests. If application of the $\widetilde{CSS1}$ test results in rejection of the no-jumps null hypothesis, then we have very strong evidence of jumps in the $DGPs$. If application of the $\widetilde{CSS1}$ does not result in rejection, then it is advisable to check this result by applying the ASJ test and CSS tests, which are more powerful.

8   Concluding Remarks

In this paper, we carry out a Monte Carlo investigation of long time span jump tests, which are designed to indicate whether the jump intensity in the underlying DGP is identically zero. The finite sample performance of these tests is compared with that of various fixed time span jump tests. We find that the long time span tests have good finite sample properties. However, we also find that fixed time span tests suffer not only from sequential bias (as is well documented), but are also severely over-sized when they are directly used to test for jumps with long time spans of data. These results confirm the findings of Huang and Tauchen (2005) that using asymptotic approximations associated with finite time span tests in order to study long time spans of data can lead to test failure. The exception to these findings is the ASJ of Aït-Sahalia and Jacod (2009), which performs favorably, when compared with Corradi et al. (2014, 2018) type long time span

---

[15]Only a small subset of our empirical findings are included in the table. These findings are illustrative of those based on examination of our complete set of findings, which are available upon request. Also note that for the CSS test, T and Δ should be carefully chosen to guarantee good finite sample properties, as shown in Section 6. Thus when our data are sampled at 5-minute frequency, we are mostly interested in the CSS test results based on the quarterly data.

jump tests. In an empirical illustration, we show that all of the jump tests that are designed to be consistent, for $T \to \infty$, find less prevalence of jumps that when fixed time span jump tests are applied using daily data.

References

Aït-Sahalia, Y. 2002a. Maximum likelihood estimation of discretely sampled diffusions: A closed-form approximation approach. *Econometrica* 70:223–262.

Aït-Sahalia, Y. 2002b. Telling from discrete data whether the underlying continuous–time model is a diffusion. *The Journal of Finance* 57:2075–2112.

Aït-Sahalia, Y., Cacho-Diaz, J., and Laeven, R. J. 2015. Modeling financial contagion using mutually exciting jump processes. *Journal of Financial Economics* 117:585–606.

Aït-Sahalia, Y., Fan, J., Laeven, R. J., Wang, C. D., and Yang, X. 2017. Estimation of the continuous and discontinuous leverage effects. *Journal of the American Statistical Association* 112:1744–1758.

Aït-Sahalia, Y. and Jacod, J. 2009. Testing for jumps in a discretely observed process. *The Annals of Statistics* 37:184–222.

Aït-Sahalia, Y. and Jacod, J. 2011. Testing whether jumps have finite or infinite activity. *The Annals of Statistics* 39:1689–1719.

Aït-Sahalia, Y., Jacod, J., and Li, J. 2012. Testing for jumps in noisy high frequency data. *Journal of Econometrics* 168:207–222.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. 2007a. Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics* 89:701–720.

Andersen, T. G., Bollerslev, T., and Dobrev, D. 2007b. No-arbitrage semi-martingale restrictions for continuous-time volatility models subject to leverage effects, jumps and iid noise: Theory and testable distributional implications. *Journal of Econometrics* 138:125–180.

Andrews, D. W. and Cheng, X. 2012. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* 80:2153–2211.

Barndorff-Neilsen, O. E. and Shephard, N. 2003. Realised power variation and stochastic volatility variance. *Bernoulli* 9:243–265.

Barndorff-Nielsen, O. E. 2002. Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 64:253–280.

Barndorff-Nielsen, O. E. and Shephard, N. 2004. Power and bipower variation with stochastic volatility and jumps. *Journal of Financial Econometrics* 2:1–37.

BARNDORFF-NIELSEN, O. E. AND SHEPHARD, N. 2006. Econometrics of testing for jumps in financial economics using bipower variation. *Journal of Financial Econometrics* 4:1–30.

BENJAMINI, Y. AND HOCHBERG, Y. 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)* 57:289–300.

BOWSHER, C. G. 2007. Modelling security market events in continuous time: Intensity based, multivariate point process models. *Journal of Econometrics* 141:876–912.

CHERNOV, M., GALLANT, A. R., GHYSELS, E., AND TAUCHEN, G. 2003. Alternative models for stock price dynamics. *Journal of Econometrics* 116:225–257.

CHRISTENSEN, K., OOMEN, R. C. A., AND PODOLSKIJ, M. 2014. Fact or friction: Jumps at ultra high frequency. *Journal of Financial Economics* 114:576–599.

CORRADI, V., SILVAPULLE, M. J., AND SWANSON, N. R. 2014. Consistent pretesting for jumps. Working Paper, Rutgers University.

CORRADI, V., SILVAPULLE, M. J., AND SWANSON, N. R. 2018. Testing for jumps and jump intensity path dependence. *Journal of Econometrics,* forthcoming.

CORSI, F., PIRINO, D., AND RENO, R. 2009. Volatility forecasting: The jumps do matter. Working Paper, University of Siena.

CORSI, F., PIRINO, D., AND RENO, R. 2010. Threshold bipower variation and the impact of jumps on volatility forecasting. *Journal of Econometrics* 159:276–288.

DUMITRU, A.-M. AND URGA, G. 2012. Identifying jumps in financial assets: A comparison between nonparametric jump tests. *Journal of Business & Economic Statistics* 30:242–255.

HOLM, S. 1979. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics* 6:65–70.

HUANG, X. AND TAUCHEN, G. 2005. The relative contribution of jumps to total price variance. *Journal of Financial Econometrics* 3:456–499.

KALNINA, I. AND XIU, D. 2017. Nonparametric estimation of the leverage effect: A trade-off between robustness and efficiency. *Journal of the American Statistical Association* 112:384–396.

LEE, S. S. AND MYKLAND, P. A. 2008. Jumps in financial markets: A new nonparametric test and jump dynamics. *Review of Financial Studies* 21:2535–2563.

MANCINI, C. 2009. Non-parametric threshold estimation for models with stochastic diffusion coefficient and jumps. *Scandinavian Journal of Statistics* 36:270–296.

PATTON, A. J. AND SHEPPARD, K. 2015. Good volatility, bad volatility: Signed jumps and the persistence of volatility. *Review of Economics and Statistics* 97:683–697.

PODOLSKIJ, M. AND VETTER, M. 2009a. Bipower-type estimation in a noisy diffusion setting. *Stochastic Processes and Their Applications* 119:2803–2831.

PODOLSKIJ, M. AND VETTER, M. 2009b. Estimation of volatility functionals in the simultaneous presence of microstructure noise and jumps. *Bernoulli* 15:634–658.

PODOLSKIJ, M. AND ZIGGEL, D. 2010. New tests for jumps in semimartingale models. *Statistical Inference for Stochastic Processes* 13:15–41.

ROMANO, J. P. AND WOLF, M. 2005. Stepwise multiple testing as formalized data snooping. *Econometrica* 73:1237–1282.

STOREY, J. D. 2003. The positive false discovery rate: a bayesian interpretation and the q-value. *The Annals of Statistics* 31:2013–2035.

THEODOSIOU, M. AND ZIKES, F. 2011. A comprehensive comparison of alternative tests for jumps in asset prices. Working Paper, Central Bank of Cyprus.

TODOROV, V. 2015. Jump activity estimation for pure-jump semimartingales via self-normalized statistics. *The Annals of Statistics* 43:1831–1864.

TODOROV, V. AND TAUCHEN, G. 2011. Volatility jumps. *Journal of Business & Economic Statistics* 29:356–371.

WHITE, H. 2000. A reality check for data snooping. *Econometrica* 68:1097–1126.

Table 1: Data Generating Processes ($DGPs$) Used in Monte Carlo Experiments

Panel A: Parameter Values

$$\kappa_\sigma, \bar{\sigma}, \zeta = \{5,\ 0.12,\ 0.5\}$$
$$\mu = 0.05$$
$$\Delta = 1/78$$
$$\rho = \{0,\ \text{-}0.5\}$$
$$\lambda = \{0.1,\ 0.4,\ 0.8\}$$

$J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(\mu_J, \sigma_J^2),$ $\{\mu_J,\ \sigma_J\} = \{0,\ 0.25\},$ $\{\mu_J,\ \sigma_J\} = \{0,\ 5 \times 0.25\}$
$\{\mu_J,\ \sigma_J\} = \{\sqrt{0.5},\ 0.25\},$ $\{\mu_J,\ \sigma_J\} = \{2.5 \times \sqrt{0.5},\ 5 \times 0.25\}$

Panel B: Data Generating Processes ($DGPs$)

DGP 1: Eq. (25) with $\mu = 0.05$, $\rho = 0$, $\kappa_\sigma = 5$, $\bar{\sigma} = 0.12$, $\zeta = 0.5$
DGP 2: Eq. (25) with $\mu = 0.05$, $\rho = $ -0.5, $\kappa_\sigma = 5$, $\bar{\sigma} = 0.12$, $\zeta = 0.5$
DGP 3: DGP 1 $+ J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 0.25^2)$
DGP 4: DGP 2 $+ J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, 0.25^2)$
DGP 5: DGP 1 $+ J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, (5 \times 0.25)^2)$
DGP 6: DGP 2 $+ J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(0, (5 \times 0.25)^2)$
DGP 7: DGP 1 $+ J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(\sqrt{0.5}, 0.25^2)$
DGP 8: DGP 2 $+ J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(\sqrt{0.5}, 0.25^2)$
DGP 9: DGP 1 $+ J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(2.5 \times \sqrt{0.5}, (5 \times 0.25)^2)$
DGP 10: DGP 2 $+ J_i \overset{\text{i.i.d}}{\sim} \mathcal{N}(2.5 \times \sqrt{0.5}, (5 \times 0.25)^2)$

*Notes: DGP 1 is a continuous process without leverage effect and DGP 2 is a continuous process with leverage effect. $DGPs$ 3-10 are continuous processes with or without leverage effect plus jumps characterized by various jump size densities. See Section 6 for complete details.

Table 2: Empirical Size of Daily Fixed Time Span Tests and Sequential Testing

Test	Subject	T = 1	T = 5	T = 50	T = 150	T = 300	T = 500
				DGP 1			
	Jump Days	0.112	0.115	0.118	0.120	0.119	0.120
		0.058	0.058	0.059	0.059	0.059	0.060
	Sequential	0.112	0.458	0.999	1.000	1.000	1.000
ASJ	Testing Bias	0.058	0.256	0.953	1.000	1.000	1.000
				DGP 2			
	Jump Days	0.100	0.116	0.121	0.120	0.120	0.120
		0.052	0.057	0.059	0.059	0.059	0.059
	Sequential	0.100	0.458	0.998	1.000	1.000	1.000
	Testing Bias	0.052	0.256	0.950	1.000	1.000	1.000
				DGP 1			
	Jump Days	0.122	0.150	0.155	0.152	0.153	0.153
		0.071	0.094	0.095	0.093	0.094	0.095
	Sequential	0.122	0.557	1.000	1.000	1.000	1.000
	Testing Bias	0.071	0.380	0.992	1.000	1.000	1.000
BNS				DGP 2			
	Jump Days	0.141	0.153	0.156	0.154	0.154	0.154
		0.098	0.094	0.096	0.095	0.095	0.095
	Sequential	0.141	0.571	1.000	1.000	1.000	1.000
	Testing Bias	0.098	0.398	0.993	1.000	1.000	1.000
				DGP 1			
	Jump Days	0.120	0.081	0.123	0.113	0.110	0.108
		0.043	0.031	0.050	0.049	0.046	0.045
	Sequential	0.120	0.347	0.999	1.000	1.000	1.000
	Testing Bias	0.043	0.146	0.921	0.999	1.000	1.000
PZ				DGP 2			
	Jump Days	0.105	0.075	0.122	0.113	0.110	0.108
		0.046	0.029	0.048	0.049	0.046	0.045
	Sequential	0.105	0.331	0.998	1.000	1.000	1.000
	Testing Bias	0.046	0.140	0.916	0.999	1.000	1.000

*Notes: Entries in this table denote rejection frequencies based on applications of *ASJ*, *BNS* and *PZ* daily fixed time span jump tests. Results for 0.1 (row 1) and 0.05 (row 2) significance levels are reported. *T* denotes the number of days for which daily fixed time span jump tests are applied. "Jump Days" shows the average percentage of detected jump days at 0.1 and 0.05 significance levels, respectively. "Sequential Testing Bias" shows probability of finding at least one jump at 0.1 and 0.05 significance levels, respectively. See Sections 5 and 6 for complete details.

Table 3: Empirical Power of Daily Fixed Time Span Tests

Jump Intensity	DGP 3	DGP 4	DGP 5	DGP 6	DGP 7	DGP 8	DGP 9	DGP 10
					ASJ			
$\lambda = 0.1$	0.127	0.124	0.143	0.135	0.160	0.160	0.186	0.185
	0.068	0.067	0.077	0.077	0.093	0.093	0.118	0.125
$\lambda = 0.4$	0.183	0.170	0.244	0.232	0.285	0.284	0.353	0.353
	0.104	0.103	0.150	0.143	0.167	0.172	0.262	0.262
$\lambda = 0.8$	0.232	0.242	0.345	0.356	0.415	0.431	0.548	0.533
	0.137	0.138	0.210	0.219	0.240	0.252	0.411	0.425
					BNS			
$\lambda = 0.1$	0.201	0.185	0.222	0.208	0.247	0.230	0.250	0.232
	0.154	0.124	0.179	0.150	0.208	0.177	0.211	0.179
$\lambda = 0.4$	0.315	0.274	0.379	0.359	0.432	0.404	0.441	0.411
	0.264	0.225	0.339	0.313	0.403	0.371	0.413	0.378
$\lambda = 0.8$	0.427	0.392	0.560	0.534	0.625	0.611	0.633	0.615
	0.377	0.336	0.526	0.499	0.600	0.583	0.612	0.588
					PZ			
$\lambda = 0.1$	0.191	0.182	0.218	0.211	0.239	0.234	0.241	0.235
	0.107	0.101	0.136	0.131	0.159	0.157	0.163	0.158
$\lambda = 0.4$	0.295	0.287	0.377	0.369	0.425	0.416	0.431	0.424
	0.217	0.210	0.311	0.298	0.366	0.352	0.372	0.360
$\lambda = 0.8$	0.424	0.397	0.560	0.548	0.634	0.624	0.640	0.628
	0.357	0.339	0.510	0.498	0.589	0.587	0.595	0.591

*Notes: See notes to Table 2. Rejection frequencies are given based on repeated daily applications of jump tests across $T = 500$ days, for each Monte Carlo replication. Thus, one can think of these experiments as reporting rejection frequencies of 500,000 tests (since $T = 500$ and there are 1000 Monte Carlo replications). Note that $DGPs$ 3–10 are all jump diffusions, i.e. with $\lambda > 0$. However, as λ is finite, jumps may not occur on some days. For more details, see Section 6.

Table 4: Empirical Size of Fixed Time Span Tests When Utilized Using Long Samples

Test	$T = 5$	$T = 25$	$T = 50$	$T = 150$	$T = 300$	$T = 500$
			DGP 1			
	0.113	0.109	0.119	0.114	0.135	0.147
ASJ	0.051	0.057	0.067	0.066	0.072	0.072
			DGP 2			
	0.106	0.131	0.148	0.136	0.132	0.145
	0.045	0.075	0.069	0.065	0.072	0.080
			DGP 1			
	0.132	0.136	0.142	0.150	0.194	0.215
BNS	0.070	0.071	0.075	0.081	0.109	0.127
			DGP 2			
	0.127	0.122	0.142	0.162	0.192	0.227
	0.081	0.065	0.080	0.095	0.104	0.132
			DGP 1			
	0.116	0.288	0.504	0.849	0.962	0.994
PZ	0.069	0.279	0.484	0.846	0.950	0.993
			DGP 2			
	0.119	0.290	0.504	0.869	0.974	0.995
	0.071	0.265	0.482	0.856	0.964	0.995

*Notes: See notes to Table 2. Entries are rejection frequencies based on a single application of the *ASJ*, *BNS* and *PZ* tests using long time span samples with T days, for each Monte Carlo replication. For all values of T, 1000 replications are run.

Table 5: Empirical Power of the *ASJ* Jump Test When Utilized Using Long Samples

Jump Intensity	*DGP* 3	*DGP* 4	*DGP* 5	*DGP* 6	*DGP* 7	*DGP* 8	*DGP* 9	*DGP* 10
				$T = 5$				
$\lambda = 0.1$	0.214	0.227	0.341	0.352	0.409	0.408	0.436	0.447
	0.151	0.169	0.285	0.286	0.345	0.351	0.395	0.405
$\lambda = 0.4$	0.498	0.482	0.750	0.731	0.838	0.846	0.876	0.872
	0.401	0.400	0.678	0.668	0.769	0.784	0.856	0.854
$\lambda = 0.8$	0.670	0.653	0.901	0.899	0.941	0.936	0.947	0.939
	0.565	0.563	0.839	0.833	0.897	0.895	0.934	0.923
				$T = 25$				
$\lambda = 0.1$	0.544	0.510	0.835	0.803	0.913	0.913	0.924	0.924
	0.480	0.454	0.812	0.779	0.907	0.900	0.921	0.918
$\lambda = 0.4$	0.908	0.898	0.993	0.996	0.992	0.991	0.986	0.984
	0.872	0.870	0.993	0.994	0.991	0.991	0.986	0.984
$\lambda = 0.8$	0.988	0.988	0.995	0.995	0.979	0.984	0.989	0.985
	0.985	0.977	0.995	0.994	0.978	0.980	0.986	0.978
				$T = 50$				
$\lambda = 0.1$	0.711	0.714	0.960	0.956	0.989	0.986	0.991	0.990
	0.663	0.655	0.954	0.949	0.985	0.985	0.990	0.990
$\lambda = 0.4$	0.987	0.989	0.997	0.998	0.992	0.996	0.993	0.994
	0.979	0.983	0.997	0.997	0.992	0.994	0.993	0.992
$\lambda = 0.8$	0.995	0.997	0.996	0.996	0.990	0.990	0.997	0.996
	0.995	0.994	0.995	0.995	0.989	0.988	0.995	0.994
				$T = 150$				
$\lambda = 0.1$	0.965	0.961	0.999	0.999	0.998	0.997	0.996	0.996
	0.954	0.947	0.999	0.999	0.997	0.997	0.996	0.996
$\lambda = 0.4$	0.998	1.000	1.000	1.000	0.998	0.999	1.000	1.000
	0.998	0.999	1.000	1.000	0.998	0.998	1.000	1.000
$\lambda = 0.8$	0.999	0.999	0.999	0.999	1.000	1.000	1.000	1.000
	0.999	0.999	0.999	0.999	1.000	1.000	1.000	1.000
				$T = 300$				
$\lambda = 0.1$	0.996	0.996	0.999	0.999	0.998	0.998	0.999	1.000
	0.995	0.995	0.999	0.999	0.998	0.998	0.999	1.000
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 500$				
$\lambda = 0.1$	0.998	0.999	0.999	0.999	0.999	0.999	1.000	1.000
	0.998	0.999	0.999	0.999	0.999	0.999	1.000	1.000
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

*Notes: See notes to Tables 3 and 4.

Table 6: Empirical Power of the *BNS* Jump Test When Utilized Using Long Samples

Jump Intensity	*DGP* 3	*DGP* 4	*DGP* 5	*DGP* 6	*DGP* 7	*DGP* 8	*DGP* 9	*DGP* 10
				$T = 5$				
$\lambda = 0.1$	0.264	0.270	0.396	0.390	0.452	0.446	0.464	0.458
	0.197	0.207	0.340	0.344	0.408	0.410	0.423	0.423
$\lambda = 0.4$	0.588	0.599	0.802	0.798	0.885	0.883	0.889	0.887
	0.533	0.540	0.771	0.768	0.871	0.872	0.880	0.878
$\lambda = 0.8$	0.817	0.815	0.948	0.961	0.983	0.981	0.986	0.984
	0.768	0.776	0.942	0.953	0.979	0.978	0.984	0.981
				$T = 25$				
$\lambda = 0.1$	0.537	0.550	0.824	0.833	0.915	0.917	0.928	0.930
	0.471	0.476	0.796	0.804	0.906	0.907	0.923	0.923
$\lambda = 0.4$	0.945	0.943	0.998	0.997	1.000	1.000	1.000	1.000
	0.923	0.928	0.998	0.997	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	0.999	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 50$				
$\lambda = 0.1$	0.707	0.720	0.950	0.958	0.987	0.988	0.993	0.995
	0.625	0.650	0.946	0.944	0.983	0.985	0.993	0.994
$\lambda = 0.4$	0.997	0.996	1.000	1.000	1.000	1.000	1.000	1.000
	0.994	0.995	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 150$				
$\lambda = 0.1$	0.947	0.958	1.000	1.000	1.000	1.000	1.000	1.000
	0.917	0.934	0.999	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 300$				
$\lambda = 0.1$	0.994	0.996	1.000	1.000	1.000	1.000	1.000	1.000
	0.993	0.994	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 500$				
$\lambda = 0.1$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

*Notes: See notes to Tables 3 and 4.

Table 7: Empirical Power of the *PZ* Jump Test When Utilized Using Long Samples

Jump Intensity	*DGP* 3	*DGP* 4	*DGP* 5	*DGP* 6	*DGP* 7	*DGP* 8	*DGP* 9	*DGP* 10
				$T = 5$				
$\lambda = 0.1$	0.316	0.317	0.418	0.413	0.465	0.456	0.471	0.464
	0.277	0.284	0.385	0.387	0.433	0.433	0.439	0.441
$\lambda = 0.4$	0.682	0.661	0.829	0.810	0.888	0.879	0.890	0.880
	0.656	0.645	0.818	0.801	0.881	0.874	0.883	0.875
$\lambda = 0.8$	0.874	0.865	0.962	0.965	0.981	0.984	0.982	0.985
	0.864	0.851	0.960	0.960	0.979	0.980	0.980	0.981
				$T = 25$				
$\lambda = 0.1$	0.788	0.777	0.911	0.900	0.936	0.935	0.939	0.939
	0.780	0.766	0.907	0.896	0.934	0.933	0.937	0.937
$\lambda = 0.4$	0.993	0.992	1.000	1.000	1.000	1.000	1.000	1.000
	0.993	0.992	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 50$				
$\lambda = 0.1$	0.960	0.954	0.991	0.987	0.997	0.993	0.998	0.994
	0.956	0.950	0.991	0.987	0.997	0.993	0.998	0.994
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 150$				
$\lambda = 0.1$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 300$				
$\lambda = 0.1$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
				$T = 500$				
$\lambda = 0.1$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.4$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
$\lambda = 0.8$	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

*Notes: See notes to Tables 3 and 4.

Table 8: Empirical Size of the $CSS1$ and $\widetilde{CSS1}$ Jump Tests

Test Statistic	Leverage	$T = 5$	$T = 25$	$T = 50$	$T = 150$	$T = 300$	$T = 500$
$CSS1$	$\varnothing$	0.202	0.151	0.129	0.123	0.106	0.112
		0.145	0.094	0.072	0.076	0.055	0.057
	$\sqrt{}$	0.264	0.316	0.461	0.858	0.987	0.999
		0.185	0.221	0.341	0.775	0.968	0.997
$\widetilde{CSS1}$	$\varnothing$	0.075	0.007	0.001	0.000	0.000	0.000
		0.048	0.003	0.000	0.000	0.000	0.000
	$\sqrt{}$	0.089	0.002	0.000	0.000	0.000	0.000
		0.053	0.001	0.000	0.000	0.000	0.000

*Notes: As in Tables 2-7, jump test rejection frequencies are reported. As discussed in Section 6, the subsampling interval, $\widetilde{\Delta}$, used in constructing critical values for the tests has been selected using a simple rule. Namely, {T=25 and 50, $\widetilde{\Delta}_n^1$=1/3, $\widetilde{\Delta}_n^2$=1/26}; {T=150, $\widetilde{\Delta}_n^1$=1, $\widetilde{\Delta}_n^2$=1/13}; {T=300, $\widetilde{\Delta}_n^1$=2, $\widetilde{\Delta}_n^2$=1/13}; {T=500, $\widetilde{\Delta}_n^1$=3.2, $\widetilde{\Delta}_n^2$=1/10}, where $\widetilde{\Delta}_n^1$ is the subsampling interval used in bootstrapping critical values for $\widetilde{CSS1}$ and $\widetilde{\Delta}_n^2$ is the subsampling interval used in bootstrapping critical values for $CSS1$.

Table 9: Empirical Size of the CSS Jump Test

		$T = 40$	$T = 50$	$T = 60$	$T = 70$	$T = 80$	$T = 90$
		No Leverage					
Δ	$\kappa = 2$	0.049	0.057	0.114	0.105	0.147	0.169
		0.019	0.032	0.057	0.060	0.079	0.103
	$\kappa = 5$	0.055	0.078	0.089	0.117	0.111	0.153
		0.025	0.042	0.048	0.053	0.059	0.097
	$\kappa = 10$	0.061	0.076	0.091	0.113	0.123	0.152
		0.021	0.033	0.042	0.067	0.067	0.089
		Leverage					
	$\kappa = 2$	0.047	0.117	0.176	0.269	0.341	0.459
		0.021	0.059	0.099	0.189	0.242	0.361
	$\kappa = 5$	0.085	0.136	0.203	0.275	0.350	0.437
		0.046	0.077	0.128	0.182	0.249	0.330
	$\kappa = 10$	0.085	0.128	0.180	0.284	0.386	0.436
		0.046	0.062	0.116	0.171	0.253	0.332
		$T = 120$	$T = 140$	$T = 160$	$T = 180$	$T = 200$	$T = 220$
		No Leverage					
	$\kappa = 2$	0.030	0.024	0.046	0.057	0.068	0.083
		0.008	0.004	0.015	0.020	0.033	0.045
	$\kappa = 5$	0.033	0.047	0.038	0.068	0.070	0.083
		0.015	0.020	0.021	0.029	0.037	0.040
	$\kappa = 10$	0.046	0.056	0.067	0.075	0.079	0.089
		0.015	0.022	0.028	0.032	0.038	0.048
$\Delta/2$		Leverage					
	$\kappa = 2$	0.041	0.075	0.123	0.160	0.234	0.270
		0.017	0.042	0.066	0.086	0.151	0.174
	$\kappa = 5$	0.062	0.087	0.127	0.170	0.227	0.290
		0.032	0.043	0.064	0.086	0.139	0.185
	$\kappa = 10$	0.071	0.101	0.156	0.167	0.246	0.283
		0.027	0.050	0.091	0.084	0.138	0.187

*Notes: Rejection frequencies are reported based on application of CSS test. Sampling intervals are either $\Delta = 1/78$ or $\Delta/2 = 1/156$. $\kappa = T^+/T$, where T^+ is the longer time span. Different combinations of Δ, T and T^+ are specified in order to find a suitable set of parameters for which adequate finite sample properties characterize the test. For complete details, see Section 6.

Table 10: Empirical Power of *CSS1* Jump Test

Jump Intensity	*DGP* 3	*DGP* 4	*DGP* 5	*DGP* 6	*DGP* 7	*DGP* 8	*DGP* 9	*DGP* 10
				$T = 5$				
$\lambda = 0.1$	0.346	0.361	0.435	0.463	0.487	0.517	0.507	0.540
	0.274	0.291	0.374	0.396	0.436	0.463	0.461	0.491
$\lambda = 0.4$	0.572	0.585	0.717	0.712	0.895	0.883	0.901	0.895
	0.498	0.514	0.665	0.650	0.879	0.865	0.890	0.887
$\lambda = 0.8$	0.664	0.677	0.790	0.784	0.968	0.973	0.969	0.976
	0.600	0.615	0.752	0.737	0.962	0.964	0.964	0.970
				$T = 25$				
$\lambda = 0.1$	0.472	0.535	0.713	0.733	0.913	0.921	0.926	0.941
	0.400	0.447	0.656	0.670	0.895	0.896	0.917	0.928
$\lambda = 0.4$	0.630	0.634	0.690	0.676	0.997	0.998	1.000	1.000
	0.560	0.537	0.604	0.598	0.996	0.997	0.998	1.000
$\lambda = 0.8$	0.650	0.645	0.647	0.650	1.000	1.000	1.000	1.000
	0.583	0.569	0.580	0.572	1.000	1.000	1.000	1.000
				$T = 50$				
$\lambda = 0.1$	0.559	0.587	0.719	0.736	0.982	0.981	0.990	0.990
	0.459	0.492	0.644	0.657	0.977	0.975	0.989	0.988
$\lambda = 0.4$	0.641	0.656	0.642	0.628	1.000	1.000	0.999	0.999
	0.557	0.561	0.548	0.559	1.000	1.000	0.999	0.999
$\lambda = 0.8$	0.623	0.604	0.628	0.627	1.000	1.000	1.000	1.000
	0.534	0.527	0.531	0.539	1.000	1.000	1.000	1.000
				$T = 150$				
$\lambda = 0.1$	0.763	0.786	0.847	0.828	1.000	1.000	1.000	1.000
	0.722	0.732	0.804	0.788	1.000	0.999	1.000	1.000
$\lambda = 0.4$	0.762	0.768	0.807	0.802	1.000	1.000	1.000	1.000
	0.716	0.717	0.759	0.764	1.000	1.000	1.000	1.000
$\lambda = 0.8$	0.761	0.750	0.805	0.795	1.000	1.000	1.000	1.000
	0.715	0.704	0.756	0.751	1.000	1.000	1.000	1.000
				$T = 300$				
$\lambda = 0.1$	0.766	0.784	0.819	0.825	1.000	1.000	1.000	1.000
	0.709	0.738	0.784	0.786	1.000	1.000	1.000	1.000
$\lambda = 0.4$	0.789	0.768	0.780	0.791	1.000	1.000	1.000	1.000
	0.742	0.713	0.733	0.755	1.000	1.000	1.000	1.000
$\lambda = 0.8$	0.759	0.756	0.808	0.802	1.000	1.000	1.000	1.000
	0.711	0.702	0.764	0.757	1.000	1.000	1.000	1.000
				$T = 500$				
$\lambda = 0.1$	0.791	0.819	0.865	0.855	1.000	1.000	1.000	1.000
	0.755	0.790	0.833	0.825	1.000	1.000	1.000	1.000
$\lambda = 0.4$	0.819	0.808	0.838	0.827	1.000	1.000	1.000	1.000
	0.771	0.774	0.804	0.795	1.000	1.000	1.000	1.000
$\lambda = 0.8$	0.786	0.776	0.823	0.836	1.000	1.000	1.000	1.000
	0.753	0.739	0.784	0.795	1.000	1.000	1.000	1.000

*Notes: See notes to Tables 3, 4 and 8.

Table 11: Empirical Power of $\widetilde{CSS1}$ Jump Test

Jump Intensity	DGP 3	DGP 4	DGP 5	DGP 6	DGP 7	DGP 8	DGP 9	DGP 10
				$T = 5$				
$\lambda = 0.1$	0.181	0.221	0.300	0.337	0.374	0.401	0.406	0.438
	0.145	0.162	0.272	0.287	0.348	0.355	0.390	0.407
$\lambda = 0.4$	0.428	0.443	0.651	0.642	0.856	0.837	0.871	0.866
	0.371	0.388	0.597	0.585	0.835	0.819	0.865	0.863
$\lambda = 0.8$	0.596	0.586	0.761	0.738	0.950	0.951	0.951	0.952
	0.515	0.513	0.717	0.710	0.943	0.944	0.948	0.948
				$T = 25$				
$\lambda = 0.1$	0.262	0.278	0.670	0.680	0.887	0.879	0.915	0.918
	0.228	0.233	0.644	0.656	0.875	0.863	0.915	0.915
$\lambda = 0.4$	0.567	0.565	0.824	0.811	0.998	0.999	0.998	0.998
	0.505	0.492	0.798	0.783	0.996	0.998	0.998	0.998
$\lambda = 0.8$	0.646	0.643	0.793	0.801	0.997	0.998	1.000	1.000
	0.602	0.579	0.766	0.766	0.997	0.996	1.000	1.000
				$T = 50$				
$\lambda = 0.1$	0.169	0.192	0.694	0.687	0.946	0.959	0.983	0.983
	0.129	0.149	0.630	0.624	0.933	0.947	0.978	0.978
$\lambda = 0.4$	0.436	0.416	0.708	0.696	0.998	0.999	0.998	0.997
	0.349	0.342	0.632	0.634	0.998	0.997	0.995	0.996
$\lambda = 0.8$	0.489	0.469	0.654	0.635	0.996	0.993	0.999	0.999
	0.411	0.386	0.584	0.586	0.995	0.990	0.997	0.997
				$T = 150$				
$\lambda = 0.1$	0.163	0.149	0.766	0.740	0.997	1.000	1.000	1.000
	0.120	0.099	0.726	0.695	0.996	0.998	1.000	1.000
$\lambda = 0.4$	0.320	0.307	0.743	0.742	1.000	1.000	1.000	1.000
	0.245	0.215	0.701	0.709	1.000	1.000	1.000	1.000
$\lambda = 0.8$	0.390	0.353	0.709	0.712	1.000	1.000	1.000	1.000
	0.311	0.286	0.659	0.657	1.000	1.000	1.000	1.000
				$T = 300$				
$\lambda = 0.1$	0.090	0.093	0.769	0.753	1.000	1.000	1.000	1.000
	0.058	0.064	0.728	0.719	1.000	1.000	1.000	1.000
$\lambda = 0.4$	0.223	0.216	0.733	0.744	1.000	1.000	1.000	1.000
	0.158	0.154	0.696	0.708	1.000	1.000	1.000	1.000
$\lambda = 0.8$	0.274	0.248	0.702	0.708	1.000	1.000	1.000	1.000
	0.199	0.188	0.659	0.646	1.000	1.000	1.000	1.000
				$T = 500$				
$\lambda = 0.1$	0.054	0.063	0.757	0.751	1.000	1.000	1.000	1.000
	0.026	0.038	0.710	0.700	1.000	1.000	1.000	1.000
$\lambda = 0.4$	0.153	0.158	0.717	0.700	1.000	1.000	1.000	1.000
	0.091	0.106	0.650	0.651	1.000	1.000	1.000	1.000
$\lambda = 0.8$	0.188	0.171	0.667	0.649	1.000	1.000	1.000	1.000
	0.126	0.129	0.607	0.585	1.000	1.000	1.000	1.000

*Notes: See notes to Tables 3, 4 and 8.

Table 12: Empirical Power of the CSS Jump Test

		$T = 40$	$T = 50$	$T = 60$	$T = 70$	$T = 80$	$T = 90$
		$\lambda = 0.1$					
Δ	DGP 3	0.450	0.473	0.514	0.571	0.633	0.617
		0.411	0.429	0.455	0.513	0.565	0.559
	DGP 4	0.465	0.485	0.550	0.647	0.652	0.702
		0.412	0.438	0.499	0.594	0.593	0.642
	DGP 5	0.821	0.872	0.922	0.952	0.965	0.975
		0.810	0.863	0.917	0.947	0.964	0.968
	DGP 6	0.829	0.876	0.928	0.953	0.974	0.978
		0.817	0.867	0.918	0.948	0.968	0.972
		$\lambda = 0.4$					
	DGP 3	0.806	0.835	0.843	0.847	0.868	0.872
		0.762	0.812	0.822	0.819	0.846	0.859
	DGP 4	0.792	0.831	0.828	0.862	0.866	0.887
		0.753	0.807	0.804	0.829	0.831	0.865
	DGP 5	0.997	0.999	1.000	1.000	1.000	1.000
		0.996	0.999	1.000	1.000	1.000	1.000
	DGP 6	0.996	0.997	0.997	1.000	1.000	1.000
		0.994	0.997	0.997	1.000	1.000	1.000
		$T = 120$	$T = 140$	$T = 160$	$T = 180$	$T = 200$	$T = 220$
		$\lambda = 0.1$					
$\Delta/2$	DGP 3	0.863	0.890	0.884	0.899	0.919	0.917
		0.838	0.872	0.862	0.889	0.905	0.894
	DGP 4	0.865	0.892	0.907	0.903	0.913	0.918
		0.835	0.869	0.895	0.885	0.898	0.903
	DGP 5	1.000	0.999	1.000	1.000	1.000	1.000
		1.000	0.999	1.000	1.000	1.000	1.000
	DGP 6	0.997	1.000	0.999	0.998	0.998	0.999
		0.997	1.000	0.998	0.998	0.998	0.999
		$\lambda = 0.4$					
	DGP 3	0.946	0.943	0.944	0.958	0.962	0.965
		0.939	0.938	0.938	0.950	0.954	0.959
	DGP 4	0.964	0.954	0.961	0.957	0.959	0.967
		0.958	0.949	0.952	0.949	0.948	0.958
	DGP 5	1.000	1.000	1.000	1.000	1.000	1.000
		1.000	1.000	1.000	1.000	1.000	1.000
	DGP 6	1.000	0.999	1.000	1.000	1.000	1.000
		1.000	0.999	1.000	1.000	1.000	1.000

*Notes: See notes to Table 9.

Table 13: *ASJ* Jump Test Results for ETFs

	2006	2007	2008	2009	2010	2011	2012	2013
SPY	3.264 (***)	1.694 (**)	0.002	2.579 (***)	5.213 (***)	0.745	0.874	1.745 (**)
XLB	5.011 (***)	2.528 (***)	1.941 (**)	3.207 (***)	4.161 (***)	0.870	2.682 (***)	0.986
XLE	0.952	4.862 (***)	0.019	7.023 (***)	1.061	0.058	5.665 (***)	1.772 (**)
XLF	3.207 (***)	4.128 (***)	1.825 (**)	1.774 (**)	0.843	0.822	1.286 (*)	1.663 (**)
XLI	4.233 (***)	5.903 (***)	1.625 (*)	1.827 (**)	7.367 (***)	1.486 (*)	0.951	1.813 (**)
XLK	7.909 (***)	4.214 (***)	0.991	1.766 (**)	1.384 (*)	0.759	0.922	2.492 (***)
XLP	3.180 (***)	8.996 (***)	7.979 (***)	4.212 (***)	0.373	1.493 (*)	6.078 (***)	1.565 (*)
XLU	7.066 (***)	2.730 (***)	1.481 (*)	10.000 (***)	0.528	0.642	2.769 (***)	3.324 (***)
XLV	7.030 (***)	6.084 (***)	1.828 (**)	2.386 (***)	2.352 (***)	1.729 (**)	0.866	2.530 (***)
XLY	3.368 (***)	1.845 (**)	3.279 (***)	4.399 (***)	0.457	0.735	0.022	2.721 (***)

*Notes: See notes to Tables 4 and 5. Entries are jump test statistics, and (***), (**), and (*) indicate rejections of the "no jump" null hypothesis at 0.01, 0.05 and 0.1 significance levels, respectively.

Table 14: *CSS1* Jump Test Results for ETFs

	2006	2007	2008	2009	2010	2011	2012	2013
SPY	2.04E-07	-3.88E-06	4.87E-04 (***)	2.65E-05	1.65E-06	-1.19E-05	-3.15E-07	-2.90E-06 (**)
XLB	-1.40E-05	-8.45E-05 (***)	1.35E-03 (***)	-1.90E-04 (***)	-9.21E-05 (***)	9.39E-06	-1.33E-06	1.87E-06
XLE	2.38E-05 (***)	-2.43E-05 (**)	1.07E-03 (**)	-9.77E-05 (**)	4.74E-05 (***)	-4.76E-05	-6.35E-06 (**)	-4.45E-06
XLF	-1.11E-05 (***)	-2.58E-04 (***)	2.00E-03 (***)	-2.26E-05	-5.04E-05 (**)	-3.69E-05	3.76E-06	-6.79E-06 (**)
XLI	-2.03E-05 (***)	9.96E-05 (***)	7.23E-04 (***)	-9.39E-05 (**)	3.24E-04 (***)	9.48E-06	1.98E-06	-2.93E-06 (**)
XLK	-1.99E-04 (***)	-6.68E-05 (***)	1.86E-03 (***)	-2.68E-05	-1.45E-04 (***)	-9.44E-06	-2.63E-06	-3.39E-06 (***)
XLP	5.92E-06 (***)	1.83E-05 (***)	-2.02E-03 (***)	-5.91E-05 (***)	-2.19E-05 (**)	-4.69E-06	-4.30E-05 (***)	8.25E-07
XLU	-9.42E-05 (***)	1.67E-04 (***)	-8.17E-05	-1.04E-01 (***)	1.75E-04 (***)	-1.91E-05	7.89E-06 (***)	-7.47E-06
XLV	-1.17E-04 (***)	9.52E-05 (***)	4.08E-04 (***)	-4.11E-05 (***)	-1.71E-04 (***)	1.24E-05	-8.03E-07	-2.53E-06
XLY	-8.05E-05 (***)	-1.06E-04 (***)	-2.97E-03 (***)	-2.72E-04 (***)	1.62E-05	-1.56E-05	-5.33E-06 (**)	2.57E-07

*Notes: See notes to Tables 8 and 13.

Table 15: $\widetilde{CSS1}$ Jump Test Results for ETFs

	2006	2007	2008	2009	2010	2011	2012	2013
SPY	1.28E-08	-2.43E-07 (*)	3.05E-05	1.66E-06	1.03E-07	-7.46E-07	-1.98E-08 (*)	-1.82E-07
XLB	-8.78E-07 (*)	-5.31E-06 (**)	8.41E-05	-1.19E-05	-5.77E-06	5.88E-07	-8.38E-08 (*)	1.17E-07
XLE	1.49E-06	-1.53E-06	6.71E-05	-6.12E-06	2.97E-06	-2.98E-06	-3.99E-07	-2.79E-07
XLF	-6.97E-07	-1.62E-05 (**)	1.25E-04	-1.42E-06	-3.16E-06	-2.31E-06	2.36E-07 (*)	-4.26E-07
XLI	-1.27E-06	6.25E-06	4.52E-05	-5.88E-06	2.03E-05 (***)	5.94E-07	1.24E-07 (*)	-1.83E-07
XLK	-1.25E-05 (***)	-4.19E-06 (**)	1.16E-04	-1.68E-06	-9.06E-06 (**)	-5.91E-07	-1.66E-07	-2.12E-07
XLP	3.71E-07 (*)	1.15E-06	-1.26E-04 (***)	-3.70E-06 (**)	-1.37E-06 (*)	-2.94E-07	-2.70E-06 (***)	5.17E-08
XLU	-5.92E-06	1.05E-05 (***)	-5.11E-06 (*)	-6.49E-03 (***)	1.10E-05 (***)	-1.19E-06	4.96E-07	-4.68E-07
XLV	-7.37E-06 (***)	5.98E-06	2.55E-05	-2.57E-06	-1.07E-05 (***)	7.76E-07	-5.05E-08 (*)	-1.58E-07
XLY	-5.05E-06	-6.68E-06 (**)	-1.86E-04 (**)	-1.71E-05	1.01E-06	-9.79E-07	-3.35E-07 (*)	1.61E-08

*Notes: See notes to Tables 8 and 13.

Table 16: *ASJ* Jump Test Results for Individual Stocks

	2006	2007	2008	2009	2010	2011	2012	2013
American Express	3.115 (***)	4.248 (***)	2.044 (**)	2.672 (***)	0.996	2.343 (***)	3.555 (***)	2.464 (***)
Bank of America	2.914 (***)	3.014 (***)	1.529 (*)	3.576 (***)	0.819	3.171 (***)	1.784 (**)	0.803
Cisco	3.811 (***)	5.997 (***)	0.180	1.818 (**)	2.865 (***)	1.665 (**)	2.581 (***)	0.922
Citigroup	3.215 (***)	0.802	0.520	3.302 (***)	6.023 (***)	0.752	0.197	0.895
Coca-Cola	8.039 (***)	10.005 (***)	6.134 (***)	4.563 (***)	0.826	1.774 (**)	2.551 (***)	3.476 (***)
Intel	2.632 (***)	4.142 (***)	3.332 (***)	0.914	5.627 (***)	6.425 (***)	1.821 (**)	1.772 (**)
JPMorgan	3.446 (***)	2.532 (***)	3.232 (***)	0.962	1.733 (**)	3.461 (***)	0.987	0.983
Merck & Co.	5.700 (***)	8.016 (***)	1.909 (**)	3.184 (***)	0.051	1.559 (*)	2.648 (***)	0.246
Microsoft	2.982 (***)	6.997 (***)	0.909	0.811	0.652	3.434 (***)	4.579 (***)	0.370
Procter & Gamble	3.285 (***)	4.933 (***)	1.814 (**)	9.998 (***)	2.387 (***)	3.218 (***)	9.003 (***)	1.745 (**)
Pfizer	2.356 (***)	4.024 (***)	0.845	0.890	7.436 (***)	1.960 (**)	1.740 (**)	2.516 (***)
Wal-Mart	0.823	4.011 (***)	1.187	1.730 (**)	0.799	2.439 (***)	0.131	3.461 (***)

*Notes: See notes to Tables 8 and 13.

Table 17: *CSS1* Jump Test Results for Individual Stocks

	2006	2007	2008	2009	2010	2011	2012	2013
American Express	1.45E-04 (***)	-8.03E-05 (***)	4.57E-03 (***)	1.30E-03 (**)	-7.80E-05	-1.34E-04 (**)	-6.61E-05 (***)	6.24E-05 (***)
Bank of America	-8.20E-04 (***)	-1.58E-04 (***)	4.16E-03 (***)	-2.26E-02 (**)	-1.29E-04	-1.39E-03 (***)	4.47E-05	-3.16E-05
Cisco	6.94E-05 (***)	-7.64E+02 (***)	9.32E-04 (***)	9.61E-05	-2.02E-04 (***)	1.31E-04 (***)	-1.85E-05	8.56E-06
Citigroup	-3.93E-05 (***)	-7.93E-05	-1.13E-02	-1.60E-02	-1.64E-03 (***)	-2.61E-04	6.52E-05 (**)	-2.26E-05
Coca-Cola	8.49E-05 (***)	5.56E-05 (***)	-2.05E-03 (***)	1.06E-04 (***)	-3.79E-05	-5.11E-06	-2.31E-05 (***)	-1.46E-05 (***)
Intel	5.55E-05 (***)	-1.49E-04 (***)	1.57E-03 (***)	2.10E-04 (**)	1.72E-04 (***)	-3.69E-05	-1.74E-05	4.64E-05 (***)
JPMorgan	4.15E-05	-1.83E-04 (***)	1.98E-03	-2.43E-04	3.88E-05	3.54E-04 (***)	7.16E-05	1.96E-05
Merck & Co.	1.39E-04 (***)	2.12E-04 (***)	-6.85E-03 (***)	-5.17E-04 (***)	3.81E-04 (***)	4.54E-06	2.80E-05 (***)	6.49E-06
Microsoft	9.93E-06 (**)	-7.46E+02 (***)	1.76E-04	6.35E-05	-4.09E-05	-1.23E-05	-6.93E-05 (***)	1.34E-05
Procter & Gamble	2.76E-05 (***)	8.05E-05 (***)	1.37E-04	-1.29E-03 (***)	2.33E-03 (***)	-2.88E-05 (***)	2.47E-05 (***)	5.72E-06
Pfizer	2.08E-04 (***)	-2.74E-03 (***)	1.11E-04	6.97E-05	3.96E-06	8.69E-05 (**)	9.01E-06	1.78E-05 (***)
Wal-Mart	6.75E-05	8.67E-05 (***)	8.60E-04 (***)	5.27E-05 (***)	-9.57E-06	3.19E-05 (**)	6.87E-06	-9.51E-06

*Notes: See notes to Tables 8 and 13.

Table 18: $\widetilde{CSS1}$ Jump Test Results for Individual Stocks

|	2006	2007	2008	2009	2010	2011	2012	2013
American Express	9.09E-06 (***)	-5.04E-06 (*)	2.86E-04	8.13E-05	-4.88E-06	-8.42E-06	-4.16E-06	3.91E-06 (**)
Bank of America	-5.15E-05 (***)	-9.94E-06 (**)	2.60E-04 (*)	-1.42E-03	-8.10E-06	-8.69E-05	2.81E-06 (**)	-1.98E-06
Cisco	4.36E-06 (*)	-4.79E+01 (***)	5.82E-05	6.02E-06	-1.27E-05 (*)	8.23E-06	-1.17E-06	5.36E-07
Citigroup	-2.47E-06	-4.98E-06 (*)	-7.08E-04 (**)	-1.00E-03	-1.03E-04 (***)	-1.64E-05	4.10E-06 (*)	-1.42E-06
Coca-Cola	5.33E-06 (***)	3.49E-06 (**)	-1.28E-04 (**)	6.64E-06	-2.37E-06 (**)	-3.20E-07	-1.45E-06 (**)	-9.15E-07
Intel	3.48E-06 (*)	-9.33E-06	9.83E-05 (**)	1.32E-05	1.08E-05 (*)	-2.31E-06	-1.10E-06	2.91E-06
JPMorgan	2.61E-06 (*)	-1.15E-05 (**)	1.24E-04	-1.52E-05	2.43E-06	2.22E-05 (*)	4.50E-06 (*)	1.23E-06
Merck & Co.	8.72E-06 (*)	1.33E-05 (***)	-4.28E-04 (***)	-3.24E-05	2.39E-05 (***)	2.84E-07	1.76E-06 (**)	4.07E-07
Microsoft	6.23E-07 (*)	-4.68E+01 (***)	1.10E-05 (*)	3.98E-06	-2.56E-06	-7.71E-07	-4.36E-06 (**)	8.42E-07
Procter & Gamble	1.73E-06 (*)	5.05E-06 (*)	8.58E-06	-8.06E-05 (***)	1.46E-04 (***)	-1.81E-06	1.55E-06 (**)	3.58E-07
Pfizer	1.30E-05 (**)	-1.72E-04 (***)	6.94E-06 (*)	4.37E-06	2.48E-07	5.44E-06	5.67E-07 (*)	1.11E-06
Wal-Mart	4.24E-06 (**)	5.44E-06	5.38E-05 (***)	3.30E-06	-6.00E-07	2.00E-06	4.32E-07 (*)	-5.96E-07

*Notes: See notes to Tables 8 and 13.

Table 19: *CSS* Jump Test Results for Individual Stocks and The Market ETF

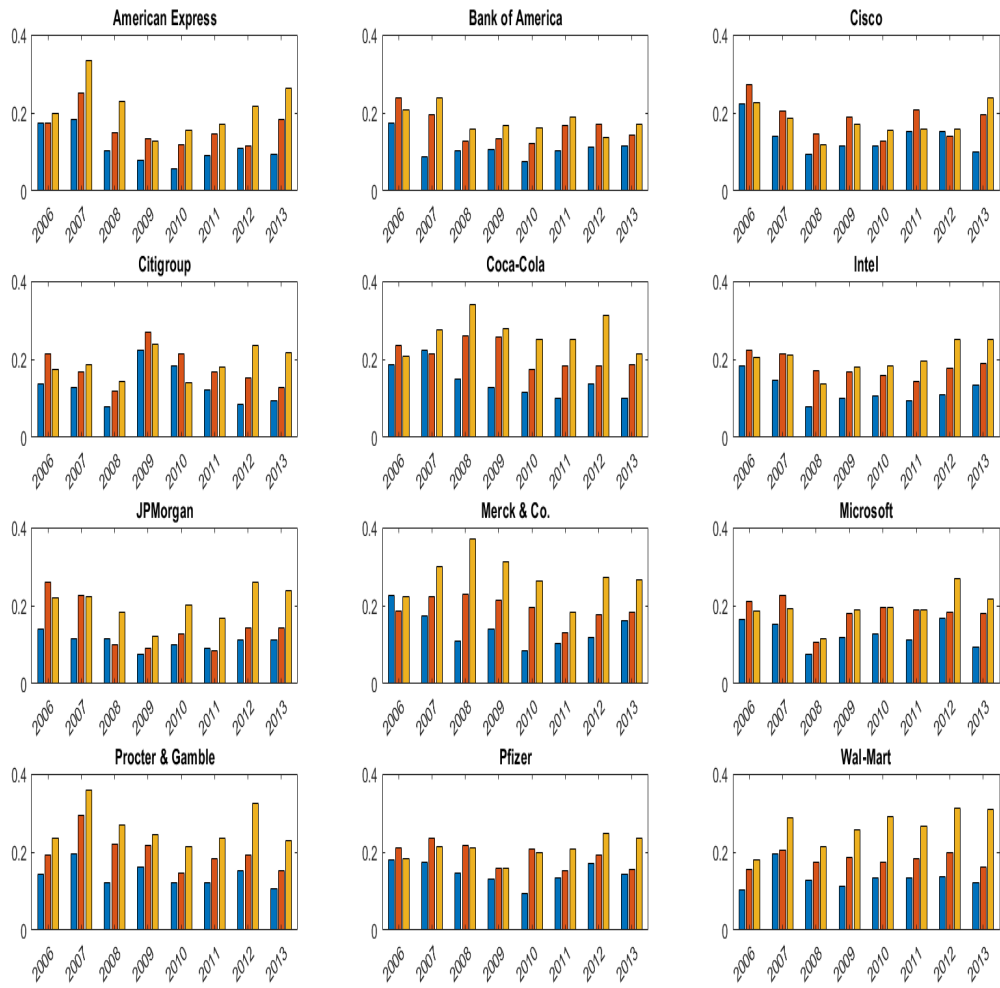|	Bank of America	Coca-Cola	Intel	JPMorgan	Merck & Co.	Pfizer	Wal-Mart	SPY
2007-Q1	10.877***	23.301***	3.148***	-2.86***	26.875***	-831.206***	16.069***	1.169
2007-Q2	-14.713***	79.895***	-1.112	-6.327***	17.705***	0.065	33.324***	-2.671***
2007-Q3	14.278***	-12.575***	9.296***	2.648***	-5.986***	-2.877***	5.97***	5.733***
2007-Q4	-7.383***	-6.614***	10.763***	-7.326***	14.327***	2.291**	4.186***	-7.235***
2008-Q1	11.617***	-7.086***	4.034***	5.499***	-279.804***	8.011***	9.386***	3.354***
2008-Q2	0.412	-65.435***	-0.965	1.722*	-3.815***	-11.948***	4.083***	-0.623
2008-Q3	3.122***	-171.387***	2.249**	1.184	-41.319***	1.016	0.694	-2.105**
2008-Q4	5.546***	-15.662***	-2.057**	4.072***	11.769***	1.692*	11.442***	8.688***
2012-Q1	1.875*	-6.077***	-2.785***	8.731***	-9.687***	-0.516	-1.131	-2.851***
2012-Q2	-2.154**	-10.595***	-4.255***	1.547	8.977***	8.513***	19.704***	0.5
2012-Q3	2.442**	-7.287***	6.378***	6.126***	26.943***	3.99***	6.503***	2.551**
2012-Q4	3.051***	-6.527***	-7.281***	-2.976***	-0.216	-7.034***	-13.85***	-0.427
2013-Q1	-5.007***	-2.223**	0.194	-1.537	-24.992***	5.515***	-11.386***	-3.096***
2013-Q2	-4.247***	-0.084	11.553***	-0.671	4.179***	0.318	-3.669***	-6.975***
2013-Q3	0.853	-1.13	-1.009	4.574***	17.83***	1.873*	3.665***	0.826
2013-Q4	8.199***	-12.131***	4.526***	11.233***	-0.141	10.736***	8.329***	6.538***

*Notes: See notes to Tables 9 and 13.

Figure 1: Annual Ratios of Jump Days for ETFs
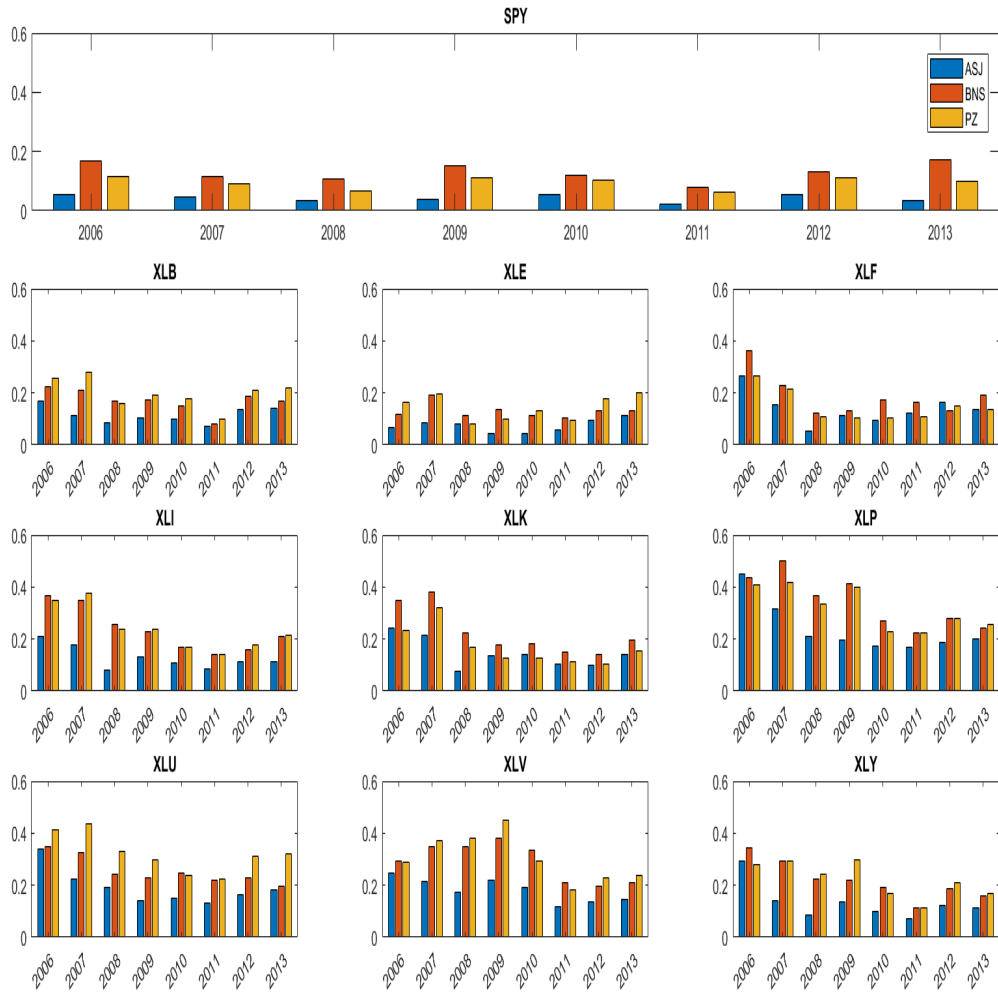
*Notes: Entries in the above charts denote annual ratios of detected jump days, based on daily applications of *ASJ*, *BNS* and *PZ* fixed time span jump tests. See Sections 5 and 7 for complete details.

Figure 2: Annual Ratios of Jump Days for Individual Stocks

*Notes: See notes to Figure 1.