Simulation-Based Exact Jump Tests in Models With Conditional Heteroskedasticity.

Lynda Khalaf,[1] Jean-Daniel Saphores,[2] and
Jean-François Bilodeau.[3]

August 8, 2002

[1]GREEN and Economics Department, Université Laval, and CRDE [Université de Montréal]. Address: Ste Foy, Québec G1K 7P4, Canada. Tel: (418) 656 2131 - 2409. Fax: (418) 656 7412. E-mail for correspondence: lkha@ecn.ulaval.ca

[2]School of Social Ecology and Economics Department, University of California, Irvine 92697. E-mail for correspondence: saphores@uci.edu

[3]Division de l'analyse et de l'élaboration de la politique de la demande, Ressources Naturelles Canada, 580 rue Booth, Ottawa, Ontario, Canada K1A 0E4.

Abstract

In models which allow for random jumps, statistical tests for jumps are typically non-standard and nuisance parameter-dependent. To handle these problems, we combine bounds and Monte-Carlo (MC) simulation techniques to derive nuisance-parameter-free bounds and obtain level-exact p-values for a wide class of processes with random jumps and time varying heteroskedasticity. When identified nuisance parameters are absent under the null, we show that MC p-values are finite sample, level-exact. To illustrate this easy-to-implement approach, we analyze the spot prices of four commodities (Aluminium, Copper, Gold and Lead) and the closing prices of four technology stocks (Intel, Microsoft, Oracle and Sun). We find significant jumps in these time series. Our approach can easily be extended to other nuisance-parameter dependent tests.

Keywords: Monte-Carlo test; bounds test; exact test; jump process; conditional heteroscedasticity.

1   Introduction

Since Merton's 1976 paper, processes with random jumps have become increasingly popular for modeling interest rates, exchange rates, stock and commodity prices, as well as natural resource prices. However, testing for jumps combines severe (and often overlooked) econometric difficulties that invalidate asymptotic methods as well as the standard bootstrap. In fact, no *provably* finite-sample jump test is currently available. In this paper, we propose a simple jump test method that guarantees type (I) error control for any sample size, for a wide class of random jump processes.

The economic literature dealing with jumps and their implications is now quite voluminous; see for example Ball and Torous (1985), Jarrow and Rosenfeld (1984), Ahn and Thompson (1988), Akgiray and Booth (1988), Jorion (1988), Brorsen and Yang (1994), Bates (1996a-b), Bakshi, Cao and Chen (1997, 2000), Drost, Nijman and Werker (1998), Bates (2000), Benzoni (2001), Pan (2001), Saphores, Khalaf, and Pelletier (2002), Das (2002), Chernov, Gallant, Ghysels and Tauchen (2002), and the references cited therein. Jump-models, which have been generalized to include mean-reversion and heteroskedasticity, are appealing because they can capture "surprise effects", i.e. large changes attributable to the arrival of unexpected information (Merton 1990). From an empirical perspective, jump models combined with processes allowing for time varying volatility, such as (G)ARCH processes, are also well-suited to capture the distributional fat-tails of many economic and financial time series. By now, it has been widely documented that economic and financial time series are compatible with both jumps and time varying volatility. However, with or without GARCH effects, statistical tests for jumps are particularly challenging for at least three reasons.

First, these tests involve nuisance parameters that are not identified under the no-jumps null hypothesis (the parameters describing the distribution of the jumps or possibly some of the GARCH parameters). In the presence of nuisance parameters, it is well known

(e.g., see Bera and Ra 1995, Hansen 1996, or Andrews 2000 and 2001) that the tests' limiting null distributions are highly non standard and, more importantly, may depend on nuisance parameters. This precludes the application of standard simulation-based size correction techniques (e.g. specialized critical point tables or the standard bootstrap).

Second, the no-jump null hypothesis sets the parameter describing the arrival of jumps at a boundary of its permissible domain (the so-called "nesting-at-boundary" problem). As demonstrated by Andrews (2001), this situation produces difficulties similar to unidentification; standard asymptotics and even bootstraps may fail.

The third difficulty stems from the GARCH parameters. These parameters are usually subject to local identifiability constraints (such as positivity and invertibility) and they intervene as nuisance parameters in the no-jump test problem. As emphasized in Dufour (1997), nuisance parameters not identified over the whole parameter space (locally almost unidentified (LAU) parameters) may cause test sizes to deviate severely from their nominal levels; both standard asymptotics and the bootstrap may thus fail. For instance, the semi-parametric GARCH(1,1) based asymptotic Wald jump test proposed by Drost, Nijman and Werker (1998) relies on the *delta*-method, and is thus not immune to LAU difficulties (Dufour 1997).

All of the above-mentioned statistical difficulties have important implications for the properties of jump tests. In particular, *spurious rejections* - resulting from test size distortions - cannot be ruled out, which underscores the importance of accounting for sample size for inference. Indeed, Dufour (1997) and Andrews (2000, 2001) show that test size distortions are not a *small sample* problem: they occur because of the *failure* of standard asymptotics rather than slow convergence. Spurious rejections may thus occur even with very large financial data sets.

Yet, in spite of the widespread application of jump models, all of the above references (with the exception of Saphores, Khalaf, and Pelletier 2002) are only justified on

2

asymptotic grounds.[1] In this context, this paper proposes a simple jump test method that guarantees type (I) error control for any sample size, for a wide class of random jump processes. More specifically, we make three contributions.

First, we show by deriving explicitly a pivotal bound that the null distribution of the test statistic we propose is bounded by a nuisance-parameter-free distribution (i.e., it is *boundedly pivotal*). This is a fundamental result in view of Dufour's (1997) *impossibility theorem*: he proves that, in LAU models, the test's size is impossible to correct unless the null distribution of a test statistic is boundedly pivotal.

Second, we rely on our pivotal bound to obtain a level-correct p-value for our jump test by applying the Monte-Carlo (MC) test technique (see Dufour 2002, Dufour and Khalaf 2001a-b). Because of the difficulties documented above, a standard bootstrap p-value, which is simulated using nuisance parameters point-estimates, is not even asymptotically valid for jump tests. In a different context, Dufour (2002) proposes to solve the nuisance parameter problem and control his MC test level by finding the maximal simulated p-value over the relevant nuisance parameter space. As this approach tends to be computationally intensive, we propose instead to find by simulation a bounds-based MC *p*-value using draws from our pivotal bound, which preserves the test level in finite samples without having to find the sup-*p*-value.[2] Furthermore, our method can easily be used for any assumed error distribution (no only Gaussian) that can be simulated.[3]

Third, to demonstrate the feasibility of our proposed test methodology, we present

---

[1]Possibly applicable asymptotic procedures (such as Hansen 1996, or Andrews 2000 and 2001) have not been shown to work for jump tests. Hansen's and Andrew's methodologies should - in principle - perform better than standard asymptotics (see for example the recent application of Hansen's methodology in Das 2002). To the best of our knowledge, however, the finite sample properties of these methods have not been studied for jump tests and these procedures have not been shown to control test size for any sample size, in contrast with our proposed methodology.

[2]Dufour and Khalaf (2001a, 2002b) use a similar approach in a different context, unrelated with jump tests.

[3]For related references on MC tests, see Dufour, Farhat, Gardiol and Khalaf (1998), Dufour, Khalaf, Bernard and Genest (2001b), Dufour and Khalaf (2001a), and Saphores, Khalaf and Pelletier (2002).

two empirical applications. We investigate the existence of jumps in the presence of GARCH(1,1) errors in the sample paths of the spot prices of four metals (Aluminium, Copper, Gold and Lead) and in the daily closing prices of four technology stocks (Intel, Microsoft, Oracle and Sun). We find ample evidence of jumps in all time series analyzed.

The paper is organized as follows. Our testing methodology is presented in Section 2. Section 3 describes our empirical applications and discusses our results. Section 4 summarizes our conclusions and offers suggestions for future research. An appendix provides basic statistical definitions, summarizes the main features of the MC test method (Dufour 2002), and documents the expressions of some of our likelihood functions.

2   Statistical Framework and Methodology

In this section, we derive a pivotal bound for a wide class of jump tests and show how to combine this bound with the MC test technique to tackle the econometric problems affecting the validity of jump tests in models with time varying volatility and random jumps.

To set focus, consider a stationary AR(1) model with GARCH (1,1) errors and Poisson jumps. If we denote the observed series by x_t, $t = 1, ..., T$, this model may be written as

$$(2.1) \qquad x_t = a_0 + a_1 x_{t-1} + \sqrt{h_t} z_t + \sum_{i=1}^{n_t} ln Y_{ti},$$

$$(2.2) \qquad h_{t+1} = \rho_0 + h_t(\rho_1 z_t^2 + \rho_2),$$

where $|a_1| < 1$ and $\rho_1 + \rho_2 \leq 1$ for stationarity, $z_t \overset{iid}{\sim} N(0,1)$, n_t is the number of jumps which occur between t and $t-1$, and Y_{ti} $(i = 1, ..., n_t)$ is the size of the ith jump in the time interval $(t-1;t)$. We also assume that jumps follow a Poisson process with arrival rate λ, and that the Y_{ti}'s are (independently) lognormally distributed with mean θ and variance δ^2. Note that n_t is an integer random variable. If $n_t = 0$, there are no jumps. If we

constrain n_t to be 0 or 1, we have a Bernoulli process. The case $a_1 = 1$, $\rho_1 = \rho_2 = 0$ yields a discretized version of Merton's Jump/Geometric-Brownian-Motion model; also setting ρ_2 to zero gives the Jump/ARCH process considered by Jorion (1988). If on the other hand a_1 is constrained to be between 0 and 1, we have a discretized Ornstein-Uhlenbeck model with jumps and GARCH(1,1) innovations.

The parameters of the mixture model (2.1)-(2.2) may be estimated by maximum likelihood. The null hypothesis under consideration is

$$(2.3) \qquad\qquad H_0 : \lambda = 0 \qquad \text{(no jump)} \,,$$

and its associated LR statistic takes the form

$$(2.4) \qquad\qquad LR = 2[L_{JAG} - L_{AG}],$$

where L_{AG} and L_{AG} are respectively the maximum of the log-likelihood function (**MLF**) associated with (2.1)-(2.2) imposing and ignoring the null hypothesis (2.3); the subscripts AG and JAG refer respectively to the (no-jumps) AR/GARCH and the Jumps/AR/GARCH models.

As recalled in the introduction, the results of Hansen (1996), Dufour (1997), and Andrews (2000, 2001) imply that the regularity conditions underlying standard asymptotics (pivotal - possibly χ^2 - limiting null distribution, the validity of the bootstrap, etc.) are not verified for LR. One reason here is that the two nuisance parameters θ and δ are not identified under H_0 (i.e., when $\lambda = 0$, the likelihood function no longer depends on θ and δ). Another reason is that the value of λ tested under H_0 is on the boundary of the parameter space. Furthermore, problems may stem from the AR/GARCH parameters a_1, ρ_1 and ρ_2 which admit LAU regions at the relevant boundaries, thus causing possible

5

discontinuities in the null limiting distributions.[4]

The approach we follow in this paper is to obtain a level-correct p-value which is *invariant* to all these nuisance parameters. Our test thus achieves level α: under the null, the largest rejection probability overall relevant nuisance parameters is $\leq \alpha$, for any sample size; see equation (A.3).

2.1 Exact jump tests

2.1.1 A bound to the null distribution of the LR jump test

Let us first introduce a bound to the LR statistic (2.4) that will allow us to easily obtain a level-exact cut-off point.

Lemma 2.1 *In the context of the jump test described by (2.1)-(2.2)-(2.3), consider the LR no-jump criterion (2.4)*

$$LR = 2[L_{JAG} - L_{AG}],$$

where L_{JAG} is the maximum of the log-likelihood function (MLF) and L_{AG} is the MLF imposing the no-jump hypothesis H_0 defined in (2.3). Let L_B denote the MLF imposing

$$H_0^B : \lambda = 0, \ a_1 = a_1^B, \ \rho_1 = \rho_1^B, \ \rho_2 = \rho_2^B,$$

where a_1^B, ρ_1^B, and ρ_2^B are known constants. If we define

$$\overline{LR} = 2\left[L_{JAG} - L_{RW}\right],$$

---

[4] The model, and test problem, are invariant to transformations of the form $p_t^* = c^* p_t + d^*$, where $c^* > 0$ and $d^* \in \Re$. It follows that LR statistics are location-scale invariant; see Dagenais and Dufour (1991). Consequently, a_0 and ρ_0 pose no particular problem here.

then $\overline{LR}$ is nuisance parameter free and it bounds LR. Indeed,

$$(2.5) \qquad\qquad \forall c, \ P(LR \geq c) \leq P(\overline{LR} \geq c).$$

As a result, LR is boundedly pivotal.

Proof. Since $H_0^B \subseteq H_0$ and since a restricted maximum is less than or equal to the unrestricted one, $L_{JAG} \geq L_{AG} \geq L_B$. This implies that

$$(2.6) \qquad\qquad [L_{JAG} - L_{AG}] \leq [L_{JAG} - L_{RW}],$$

and thus $LR \leq \overline{LR}$. Furthermore, as H_0^B restricts all the *identified* intervening parameters to known values, the null distribution of $\overline{LR}$ is nuisance parameter free.[5] The result follows (see A.2).

The statistical strategy underlying our bound $\overline{LR}$ may be summarized as follows (see Section A in the appendix). We introduce a hypothesis $[H_0^B]$ formulated so that: (i) it is a special case of the restrictions to be tested, and (ii) its associated LR criterion is pivotal. For example, the hypothesis which sets all relevant intervening parameters to known values achieves this requirement. The LR criterion associated with H_0^B [namely $\overline{LR}$] provides the desired bound. Indeed, since H_0^B is constructed as a special case of the tested hypothesis, $\overline{LR}$ is clearly larger than LR, and thus the null distribution of $\overline{LR}$ yields an upper bound applicable to LR. Most importantly, the pivotality of $\overline{LR}$ guarantees the validity of the bounds test since the cut-off point associated with the pivotal bounding statistic is *conservative*: if the observed LR exceeds the bounds-cut off point (the critical point associated with $\overline{LR}$), then the test based on LR is most certainly significant.[6]

---

[5]As mentioned above, the no-jump LR statistic is location-scale invariant. Consequently, a_0 and ρ_0 are not nuisance parameters here.

[6]For a different (although related) problem, Dufour (1997, Theorem 5.1) uses a similar strategy to bound the null distribution of the LR statistics in instrumental variable (IV) regressions with possibly

Let us now show how to obtain a nuisance-parameter-free cut-off (or alternatively a p-value) through the bounding statistic $\overline{LR}$, using MC tests.

2.1.2 Exact Monte Carlo LR jump tests

Let us first consider Merton's basic model, i.e. (2.1)-(2.2) with $a_1 = 1$, $\rho_1 = \rho_2 = 0$. Let LR_0 denote the value of the jump statistic computed from the observed data. To obtain an exact Monte Carlo jump test, simulate N samples from the no-jump data generating process where a_0 and ρ_0 are set to their constrained MLE values (imposing the no-jump null hypothesis). For all simulated samples, compute the corresponding jump test statistics, denoted LR_1, ..., LR_N. Next, calculate

$$(2.7) \qquad \widehat{p}_N(LR_0) = \frac{N\widehat{G}_N(LR) + 1}{N + 1}, \ \widehat{G}_N(LR_0) = \frac{\sum_{i=1}^{N} I_{[0,\infty]}(LR_i - LR_0)}{N},$$

where $I_{[0,\infty]}(x) = 1$ if $x \geq 0$ and $I_{[0,\infty]}(x) = 0$ otherwise. Then for all $0 < \alpha < 1$, the critical region

$$(2.8) \qquad\qquad\qquad \widehat{p}_N(LR_0) \leq \alpha$$

has exactly size α if $\alpha(N + 1)$ is an integer, in the sense that

$$(2.9) \qquad\qquad\qquad P_{(H_0)}[\widehat{p}_N(LR_0) \leq \alpha] = \alpha,$$

where $P_{(H_0)}$ refers to the probability imposing the null hypothesis.

Property (2.9) is based on a fundamental distributional result concerning the ranks of a set of exchangeable random variables (Dufour 2002). Lemma B.1 (see the appendix)

---

weak instruments. The IV-regressions test relates to the jump test we study in this paper through the identification difficulties both tests raise: poor instruments are indeed LAU problems. Of course, the jump test case is further complicated by the unidentified jump parameters.

applies here on observing that: (i) $N - N\widehat{G}_N(LR_0) + 1$ gives the rank of LR_0 in the series $LR_0, LR_1, ..., LR_N$, and (ii) LR_0 and LR_j, $j = 1$, $...$, N are exchangeable (by construction). Indeed, since the test problem is location-scale invariant, the null distribution of the simulated statistic does not depend on a_0 or ρ_0, and so the LR_j, $j = 1$, $...$, N are i.i.d. random variables with the same distribution as LR_0 under the no-jump null hypothesis. This implies that the MC p-value (2.7) is invariant to the values of a_0 and ρ_0.

As a result, the MC p-value (2.7) does not depend on the "problematic" parameters θ and δ^2. Furthermore, the boundary restriction does not intervene since the conditions underlying Lemma B.1 (no unknown parameter needs to be dealt with in order to generate $LR_1, ..., LR_N$) are satified.

Now suppose instead that the values of a_1, ρ_1, and ρ_2 are not set by the model, which is typically the case. To better understand our approach, let us see how we would proceed with a (standard) parametric bootstrap. Let LR_0 denote the observed value of (2.4). First, we would draw N simulated samples setting all parameters, namely a_0 and ρ_0, and a_1, ρ_1, and ρ_2, to their constrained MLE values $\widehat{a}_0$, $\widehat{a}_1$, $\widehat{\rho}_0$, $\widehat{\rho}_1$, and $\widehat{\rho}_2$ estimated from the data. We would then calculate (2.4) for these simulated samples and get N replications of the test statistic. Finally, we would calculate a p-value from 2.7, but this p-value would depend on the choice of a_1, ρ_1, and ρ_2, so we would denote it by

$$(2.10) \qquad \widehat{p}_N(LR_0|\widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2).$$

Since the null distribution of the simulated statistic is not pivotal, nothing guarantees that the level property $P_{(H_0)}\left[\widehat{p}_N(LR_0|\widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2) \leq \alpha\right] \leq \alpha$ holds. Furthermore, nothing guarantees regularity conditions which ensure asymptotic validity, in the sense that

$$(2.11) \qquad \lim_{T \to \infty} \left\{ P[\widehat{p}_N(LR_0|\widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2) \leq \alpha] - P[\widehat{p}_N(LR_0|a_1^0, \rho_1^0, \rho_2^0) \leq \alpha] \right\} = 0.$$

9

In (2.11), $\widehat{p}_N(LR_0|a_1^0, \rho_1^0, \rho_2^0)$ is the empirical p-value one would obtain for the "true" nuisance parameters values (which are, of course, unknown). In fact, the results of Dufour (1997, 2002) and Andrews (2000, 2001) imply that the conditions underlying (2.11) fail for the same reason that standard asymptotics fail in this context. In practice, this means that a test based on (2.10) may be spurious *even in large samples.*

To avoid these problems, one alternative to the parametric bootstrap (see Section B in the appendix) is to conduct a *maximized* Monte Carlo (MMC) test, which consists in taking the sup of the bootstrap p-values over the relevant nuisance parameters space:

$$(2.12) \qquad \sup_{a_1, \rho_1, \rho_2} [\widehat{p}_N(LR_0|a_1, \rho_1, \rho_2)] \le \alpha.$$

Such a test is level-correct by construction, in the sense that

$$(2.13) \qquad P_{(H_0)} \left[\sup_{a_1, \rho_1, \rho_2} [\widehat{p}_N(LR_0|a_1, \rho_1, \rho_2)] \le \alpha \right] \le \alpha.$$

Of course, the MMC test is only useful if the sup-p-value is non-trivial. Lemma (2.1) establishes this property for $\sup \widehat{p}_N(LR_0|a_1, \rho_1, \rho_2)$, so the MMC test is applicable. Calculating the MMC sup-p-value is computationally intensive, however, so we pursue a bounds MC (BMC) test instead.

Proposition 2.2 *In the context of the jump test described by (2.1)-(2.2) and the null hypothesis (2.3)* $\lambda = 0$, *consider the LR test statistic (2.4)*

$$LR = 2[L_{JAG} - L_{AG}]$$

where L_{AG} and L_{JAG} are the MLF imposing and ignoring $\lambda = 0$. Let LR_0 denote the value of LR computed from the observed data. Obtain simulated samples from the data generating process (2.1) and (2.2) without jumps where all parameters are set to their

10

constrained MLE values $\widehat{a}_0$, $\widehat{a}_1$, $\widehat{\rho}_0$, $\widehat{\rho}_1$, and $\widehat{\rho}_2$. For each simulated sample, compute

$$(2.14) \qquad\qquad \overline{LR}_i = 2\left[L^i_{JAG} - L^i_{RW}\right], \ \ i = 1, ..., N,$$

where L^i_{JAG} is the unconstrained MLF associated with (2.1)-(2.2) and simulated sample i; and L^i_{RW} denotes the corresponding MLF imposing

$$H^B_0 : \lambda = 0, \ a_1 = \widehat{a}_1, \ \rho_1 = \widehat{\rho}_1, \ \rho_2 = \widehat{\rho}_2.$$

In the above, $\widehat{a}_1$, $\widehat{\rho}_1$, and $\widehat{\rho}_2$ are the MLE values computed from the observed data and used to generate the simulated samples. Let

$$\widetilde{p}_N(LR_0) = \frac{N\widetilde{G}_N(LR_0) + 1}{N+1}, \quad \widetilde{G}_N(LR_0) = \frac{\sum_{i=1}^{N} I_{[0,\infty]}(\overline{LR}_i - LR_0)}{N}$$

with $I_{[0,\infty]}(x) = 1$ if $x \geq 0$ and $I_{[0,\infty]}(x) = 0$ otherwise. Then for all $0 < \alpha < 1$, the critical region

$$(2.15) \qquad\qquad \widetilde{p}_N(LR_0) \leq \alpha$$

has exactly level α if $\alpha(N+1)$ is an integer, in the sense that

$$(2.16) \qquad\qquad P_{(H_0)}\left[\widetilde{p}_N(LR_0) \leq \alpha\right] \leq \alpha$$

where $P_{(H_0)}$ refers to the probability imposing the null hypothesis.

Inequality (2.16) obtains from Lemma 2.1 with $a^B_1 = \widehat{a}_1$, $\rho^B_1 = \widehat{\rho}_1$, $\rho^B_2 = \widehat{\rho}_2$ and Lemma B.1. Indeed, the N random variables $\overline{LR}_j$, $j = 1, \ldots, N$ are independent realizations from a distribution which bounds the distribution of LR_0, in the sense of (A.5), under the null hypothesis of interest. Then $\widetilde{p}_N(LR_0)$ bounds the empirical probability to observe a

11

value as extreme or more extreme than LR_0 under the null hypothesis. Formally, (2.5) in Lemma 2.1 implies that

$$\left\{ P_{(H_0)} \left[\widetilde{p}_N(LR_0) \leq \alpha \right] \right\} \leq \left\{ P_{(H_0)} \left[\sup_{a_1, \rho_1, \rho_2} \left[\widehat{p}_N(LR_0 | a_1, \rho_1, \rho_2) \right] \leq \alpha \right] \right\}.$$

Applying (2.13) to the preceding inequality proves (2.16). The key here is, again, that no unknown parameter (including θ and δ^2) needs to be dealt with to draw realizations from the bounding statistic (??). This also takes care of the boundary restriction since the level control conditions (2.16) depend only on the availability of i.i.d. realizations from the bounding statistic.

The result of Proposition 2.2 is not limited to the bounding constraint (H_0^B in Lemma 2.1) based on $a_1^B = \widehat{a}_1$, $\rho_1^B = \widehat{\rho}_1$, $\rho_2^B = \widehat{\rho}_2$. In fact, a valid bound only requires to restrict a_1, ρ_1 and ρ_2 to (any) known values (e.g., $a_1 = 1$, $\rho_1 = \rho_2 = 0$), so long as the simulated samples are generated with these values. Furthermore, the choice of values for (a_0, ρ_0) is also irrelevant, due to the location-scale invariance of the testing problem. Note that the MMC methodology may be viewed as a numerical search for the optimal bound, i.e. the value of a_1, ρ_1 and ρ_2 which will yield the tightest bound. Of course, there is no use to search for the tightest bound if the test is significant for a valid bound.

When implementing the BMC procedure in conjunction with a MMC test, it is useful to realize that bootstrap non-rejections are *exactly* conclusive in the sense that:

$$\left\{ \widehat{p}_N(LR_0 | \widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2) > \alpha \right\} \Rightarrow \left\{ \sup_{a_1, \rho_1, \rho_2} \left[\widehat{p}_N(S_0 | a_1, \rho_1, \rho_2) \right] > \alpha \right\}.$$

We thus recommend the following three-step MC procedure, which bears similarities with the well known Durbin-Watson test:

- Obtain a bounds p-value as described in Proposition 2.2. If the bounds p-value $\leq \alpha$, then the test is significant.

12

- If the bounds p-value $> \alpha$, obtain a (standard) parametric bootstrap type p-value as explained above. If this bootstrap p-value exceeds α, then the test is not significant.

- If the bounds p-value $> \alpha$ whereas the bootstrap p-value $\leq \alpha$, derive the MMC p-value as in (2.12).

To conclude this section, it is worth emphasizing that our methodology can easily be extended to a wide class of models.

First, with respect to the GARCH component, our testing strategy is valid for the class of augmented GARCH(p,q) processes (see Duan 1997). This includes many popular processes such as the LGARCH(p,q), MGARCH(p,q), and EGARCH(p,q) (see Duan 1997 and references therein).[7] To apply the BMC methodology, realizations from the null distribution of a bounding statistic based on any null hypothesis that sets all nuisance parameters to known values could be used. The simplest possibility is the driftless random walk, which guarantees the existence of a non-trivial pivotal bounding statistic. In turn, the existence of this bound guarantees that the MMC algorithm will not trivially converge to 1. Thus, a LR-based three-step MC test as described above is *provably* immune to LAU difficulties in finite samples, a property not yet established for Wald type jump tests (e.g. Drost, Nijman and Werker 1998).

Second, the normality assumption (normality of z_t in (2.1)-(2.2)) is not necessary for the validity of our approach. For many empirical applications, heavy tailed (such as student-t- based) or asymmetric distributions may also be considered. Indeed, our results simply rely on the fact that a constrained MLF is always less than or equal to its unconstrained counterpart, whether the likelihood is Gaussian or not.

---

[7]Using Duan (1997)'s GARCH specification is relevant because this GARCH class converges [under Gaussian fundamentals] to jumpless processes as the frequency of observations increases. This is desirable, in a mixtures context, to improve identifiability of the discrete jump component; see Drost, Nijman and Worker (1998) for more insight on alternative GARCH processes with inherent jumps.

3   Empirical illustration

To illustrate our methodology, we consider two empirical applications where we test for jumps with and without GARCH effects.

3.1   Empirical Data

In the first application, we analyze the daily and weekly London Metal Exchange (LME) spot prices of four metals: Aluminium, Copper, Gold, and Lead. In the second application, we look at daily and weekly closing prices of four NASDAQ technology stocks: Intel, Microsoft, Oracle, and Sun. It is well known that the presence of jumps in time series of commodities or stocks can have serious implications on pricing futures contracts as well as derivatives (e.g., see Jorion 1988).

In both applications, our price series extend from the beginning of 1989 to the end of January 2002. The number of daily data points differs slightly between the two applications (3302 for metals and 3414 for technological stocks) because the LME sometimes closes on days where the NASDAQ is open. Weekly series (683 points) are constructed from daily data by taking the Wednesday price to avoid beginning or end of the week effects. In the rare instances where the Wednesday price is missing, we use the Tuesday price instead.

Tables 1 and 2 present summary statistics for the logarithm of the spot prices of the eight series considered. For metals (Table 1), the skewness, excess kurtosis, and Jarques-Bera statistics are significant at 1%, except for the skewness of Copper and the excess kurtosis of Aluminium. For technological stocks (Table 2), the skewness, excess kurtosis, and Jarques-Bera statistics are also significant at 1%, with the exception of the skewness of Intel and Oracle. This evidence of fat tails motivates our jump models.

14

3.2 Models

For metal spot prices, we fit to the logarithm of each time series a simple discretized Ornstein-Uhlenbeck process (OU) with and without normally distributed Bernoulli jumps and GARCH(1,1) errors. The full model can be written

$$(3.1) \qquad x_t = \mu(1 - e^{-\kappa}) + e^{-\kappa}x_{t-1} + \sqrt{h_t}z_t + \delta_{1n_t}\ln Y_t,$$

where P_t is the spot price at time t; $x_t = Ln(P_t)$; $h_{t+1} = \rho_0 + h_t(\rho_1 z_t^2 + \rho_2)$; $z_t \overset{iid}{\sim} N(0,1)$; and n_t equals one if a jump occurs between t and $t-1$ or zero otherwise, so $\delta_{1n_t} = 1$ if $n_t = 1$ and $\delta_{1n_t} = 0$ otherwise. We denote by λ the probability of arrival of a jump in a unit time interval , and assume that the jump size Y_t is (independently) lognormally distributed with mean θ and variance δ^2. We suppose that jumps follow a Bernoulli instead of a Poisson process to simplify the estimation and the interpretation of our models. We get (3.1) from (2.1)-(2.2) by setting $a_0 = \mu(1 - e^{-\kappa})$ and $a_1 = e^{-\kappa}$, where $\kappa > 0$. If there are no ARCH effects, we simply set $\rho_1 = \rho_2 = 0$, and we designate $\rho_0 > 0$ by σ^2.

By contrast, we model technological stock prices using a geometric Brownian motion (GBM) with and without normally distributed Bernoulli jumps and GARCH(1,1) errors. Using the same notations, the full model can be written

$$(3.2) \qquad x_t - x_{t-1} = \mu + \sqrt{h_t}z_t + \delta_{1n_t}\ln Y_t.$$

(3.2) is obtained from (2.1)-(2.2) by setting $a_0 = \mu$ and $a_1 = 1$. Duan (1997) shows that the diffusion limits of the no-jump components of (3.1) and (3.2) are stochastic volatility models.

While the GBM is often the starting points of stock pricing models because of its

15

attractive properties, there are theoretical reasons for modeling the spot price of commodities with a mean-reverting process. First, the GBM implies that the volatility of spot prices increases without bounds as the time horizon increases. Second, as recalled for example in Schwartz (1997), when the price of a commodity is relatively high, supply tends to increase as higher cost producers enter the market, which ends up lowering prices. Conversely, when prices are relatively low, supply over time tends to decrease as higher cost producers exit the market, which puts upward pressure on prices.

We estimate our models by maximum likelihood using the software package Gauss. The expressions of the likelihood function for (3.1), which we did not find elsewhere in the literature, is provided in Section C in the appendix; for an expression of the likelihood function for (3.2), see Ball and Torous (1985). As remarked by Ball and Torous (1985), the likelihood functions of jump-diffusion models usually have a local maximum at $\lambda = 0$ (the no jump case). To guard against numerical maximization problems, we use the procedure OPTMUM in GAUSS with multiple starting points of likely parameter values for each iteration (up to 42 for the GARCH model) for observed and simulated bootstrap maximum likelihood functions.

3.3   Results

Results are presented in Tables 3 and 4. We report estimates and standard errors, the LR test statistic (2.4) and its bootstrap and bounds p-values $\widehat{p}_N(LR_0|\widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2)$, $\widetilde{p}_N(LR_0)$ defined by (2.10) and (2.15) respectively; see Proposition 2.2.

Table 3 presents our results for the weekly metals data; a summary of the results for daily data [the estimated values of λ, the associated LR jump test statistic and its MC bound p-values] is reported in a footnote. Table 4 reports our results for the daily technological stocks data; a summary of the results for weekly data can be found in a

16

footnote.[8]

From Tables 3 and 4, we first observe that jumps are statistically significant in each of the time series investigated, with and without GARCH effects, at both daily and weekly frequencies. Indeed, both the bootstrap and bound-based p-values are 0.01 for 100 replications for all time series and all models considered, except for Lead where the bound-based p-value is 0.02.[9] There is thus no need to run the MMC test. These results illustrate the usefulness of our bound.

For the metals data, estimated daily jump frequencies (λ) with GARCH errors, range from 0.089 (one jump every $\approx$ 11 days) for Aluminium to 0.330 (one jump every $\approx$ 3 days) for Gold. For the stock series, estimated jump frequencies in the presence of GARCH errors range between 0.055 (one jump every $\approx$ 18 days) for Microsoft and 0.102 (one jump every $\approx$ 10 days) for Sun. Overall, the estimated λ is usually smaller in models with GARCH errors; Lead (at weekly frequencies) and Gold (at daily frequencies) are exceptions.

Serious caution must be exercised in comparing the values of λ for different frequencies; indeed, daily frequencies may for example reflect beginning and end of the week effects or peculiarities of the market micro-structure that are not present at lower frequencies (for weekly, monthly, or quarterly data). For more on the pitfalls of comparing jump frequency estimates, see for example the discussion in Drost, Nijman and Werker (1998) on time aggregation of jump/GARCH models. Here we simply note that jumps are significant for all the daily and weekly series we examined, and in some cases (e.g., Microsoft and

---

[8]Following Schwartz (1997) who analyzes the stochastic behavior of commodity prices using futures data (although without allowance for jumps), we focus on weekly data for metals. Conversely, empirical financial models usually rely on daily data. We report results based on daily data for metals and on weekly data for technological stocks only for completeness.

[9]No asymptotics on N is used to derive the key properties of the MC method. In fact a value as small as 19 is enough to control size; of course, power increases with N, yet in the literature on MC tests (see the references above), it is demonstrated that 99 replications is a quite reasonable choice. See also the section on Monte-Carlo tests in the appendix.

Oracle), estimated jump frequencies are higher at weekly than at daily frequencies. As recommended in the recent econometrics literature (see for example Dufour 1997, Staiger and Stock 1997, Wang and Zivot 1998), a confidence interval for λ should be based on a set estimate obtained by inverting exact tests rather than on Wald-type confidence intervals based on asymptotic standard errors; indeed, the statistical difficulties discussed above cast serious doubts on the reliability of the latter. Our exact test on λ may thus serve as a basis for a more valid set estimation method, but this is left for future work.

Interestingly, on examining GARCH coefficients, conditional heteroskedasticity seems prevalent with and without jumps. Our results suggest that neither Jump nor GARCH models can solely account for the observed fat tails in the series analyzed. Note that for Gold, although ARCH effects seem present in the GARCH-OU models (with and without jumps), the GARCH parameter does not appear to differ from zero. This observation is compatible with the findings of Schwartz (1997).

Our analyses of the metals data further reveal that the mean reversion coefficient (denoted by κ), becomes even smaller when jumps are added, with and without GARCH. For instance, in the case of the OU Aluminium model, the autoregressive coefficient increases from $\exp(-.021) = 0.979$ to $\exp(-.0075) = 0.993$ in the jump model. This observation should motivate further work on unit root test which formally account for jumps.

Finally, although we assume Gaussian errors (in (3.1)-(3.2), $z_t \overset{iid}{\sim} N(0,1)$) as Gaussian-based models often serve as useful and popular fundamental starting points, our test strategy allows for non-Gaussian fundamentals (such as t-distributed errors) without any conceptual change. Our purpose here is mostly to illustrate our testing procedure, keeping in mind that, with the exception of the jump and ARCH tests in Saphores, Khalaf, and Pelletier (2002), none of the available evidence in favor of jumps is exact, even in Gaussian contexts.

4 Conclusions

When nuisance parameters are unidentified under the null, conventional asymptotics and the standard bootstrap fail even for large samples. This is typically the case for available tests for jumps in jump models combined with processes allowing for time varying volatility such as (G)ARCH. In this paper, we propose a methodology combining boundedly pivotal statistics with the MC test technique to tackle the difficulties plaguing available jump tests. First, we establish *analytically* that the LR no-jump test statistic is boundedly pivotal. Then, we apply simulation-based methods to the LR and to bounding statistics to obtain level-exact p-values in finite samples. Our approach is simple to implement, it is applicable to a large class of parametric models, and it can easily be generalized to other test problems where unidentified nuisance parameters are present.

To illustrate the usefulness of our methodology, we conduct jump tests on the spot prices of four metals and on the closing prices of four technology stocks. We fit a mean-reverting process with normally distributed Bernoulli jumps, with and without a GARCH(1,1) errors to the logarithm of the metal price series; for the technological stocks, we estimate a geometric Brownian motion with normally distributed Bernoulli jumps, with and without a GARCH(1,1) errors. We find statistically significant jumps for all time series considered for both daily and weekly frequencies.

Our results suggest several promising avenues for further work, including the development of mean reversion tests in the presence of jumps, the construction of exact confidence sets for the jump frequency parameter, and empirical applications of our testing strategy to other GARCH (including GARCH-t or other non-Gaussian) models.

Acknowledgments

Table 1: Summary Statistics of the Logarithm of Weekly Spot Prices for Metals

	Aluminium	Copper	Gold	Lead
Mean	7.292	7.663	5.823	6.337
Standard deviation	0.165	0.227	0.140	0.209
Skewness	0.338	-0.078	-0.518	0.369
Excess kurtosis	0.032	-1.003	-1.178	-0.358
Jarques-Bera	13.019	29.326	70.001	19.135

The metals data consists of weekly Wednesday spot prices from the London Metals Exchange(LME). There are 683 observations; they cover the period extending from 01/04/89 until 01/30/02. Following Campbell, Lo and MacKinlay (1997, pages 18-20), the distributions of skewness and excess kurtosis under normality may be respectively approximated as $N(0,6/T)$ and $N(0,24/T)$, where T is the sample size. The approximate distribution for Jarques-Bera is $\chi^2(2)$. For a sample size of 683, the corresponding 1% critical points for skewness, excess kurtosis, and Jarques-Bera are thus ± 0.241, ± 0.482, and 9.21. We see that the skewness, excess kurtosis, and Jarques-Bera statistics reported in Table 1 are significant at 1%, except for the skewness of Copper and the excess kurtosis of Aluminium.

Table 2: Summary Statistics of the Logarithm of Daily Spot Prices for Technological Stocks

	Intel	Microsoft	Oracle	Sun
Mean	1.924	2.461	0.839	1.018
Standard deviation	1.347	1.456	1.537	1.413
Skewness	-0.005	-0.136	0.063	0.640
Excess kurtosis	-1.388	-1.131	-0.998	-0.874
Jarques-Bera	273.99	192.52	143.93	341.68

The technological stocks data consists of daily closing prices from the NASDAQ. There are 3414 observations; they cover the period extending from 01/02/89 until 01/31/02. Following Campbell, Lo and MacKinlay (1997, pages 18-20), the distributions of skewness and excess kurtosis under normality may be approximated as $N(0,6/T)$ and $N(0,24/T)$, respectively. The approximate distribution for Jarques-Bera is $\chi^2(2)$. For a sample size of 3414, the corresponding 1% critical points for skewness, excess kurtosis, and Jarques-Bera are thus ± 0.108, ± 0.216, and 9.21. We see that the skewness, excess kurtosis, and Jarques-Bera statistics reported in Table 1 are significant at 1%, with the exception of skewness for Intel and Oracle.

Table 3: Parameter Estimates and Bernoulli Jump Tests for Metal Prices[10]

	Aluminium		Copper	
	No Jumps	**Jumps**	**No Jumps**	**Jumps**
OU				
μ	7.249 (0.048)	7.246 (0.123)	7.549 (0.118)	7.453 (0.187)
κ	0.023 (6.5E-3)	7.9E-3 (6.2E-3)	0.011 (5.2E-3)	7.7E-3 (4.7E-3)
σ	0.028 (7.5E-4)	0.022 (9.9E-4)	0.030 (8.2E-4)	0.022 (1.7E-3)
λ		0.063 (0.029)		0.201 (0.076)
θ		-9.8E-3 (0.013)		1.8E-3 (6.0E-3)
δ		0.064 (0.013)		0.046 (6.4E-3)
MLE	1488.05	1535.79	1417.72	1451.20
LR	95.49 [0.01, 0.01]		66.97 [0.01, 0.01]	
OU GARCH				
μ	7.227 (0.046)	7.255 (0.136)	7.519 (0.167)	7.310 (0.391)
κ	0.021 (6.8E-3)	7.5E-3 (6.3E-3)	7.3E-3 (5.1E-3)	4.8E-3 (4.7E-3)
ρ_0	5.2E-5 (1.3E-5)	1.2E-5 (1.0E-5)	1.7E-4 (6.2E-5)	7.9E-5 (3.8E-5)
ρ_1	0.060 (0.019)	1.7E-3 (2.1E-3)	0.108 (0.035)	0.089 (0.036)
ρ_2	0.855 (0.029)	0.970 (0.025)	0.692 (0.093)	0.720 (0.088)
λ		0.050 (0.033)		0.167 (0.090)
θ		-7.9E-3 (0.020)		4.4E-3 (6.9E-3)
δ		0.065 (0.017)		0.040 (7.3E-3)
MLE	1523.70	1551.56	1449.29	1468.19
LR	55.71 [0.01, 0.01]		37.79 [0.01, 0.01]	

---

[10]Standard errors are in parenthesis. For the no-GARCH model, we report the bootstrap p-value $\widehat{p}_N(LR_0|\widehat{a}_1)$ [(2.10)] and the bounds p-value $\widetilde{p}_N(LR_0)$ [(2.15)]; for the GARCH-MRM, we report the bootstrap p-value $\widehat{p}_N(LR_0|\widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2)$ [(2.10)] $\widetilde{p}_N(LR_0)$ and the bound p-values $\widetilde{p}_N(LR_0)$ [(2.15)] respectively. Results with daily data for the Jump/MRM model (no-GARCH) may be summarized as follows. Aluminium: estimated $\lambda = 0.197$ and $LR = 446.004$ (MC p-value = 0.01); Copper: estimated $\lambda = 0.201$ and $LR = 416.560$ (MC p-value = 0.01); Gold: estimated $\lambda = 0.219$ and $LR = 945.366$ (MC p-value = 0.01); Lead: estimated $\lambda = 0.105$ and $LR = 636.870$ (MC p-value 0.01). The MC p-values are based on (2.8).

Table 3 (Continued): Parameter Estimates and Bernoulli Jump Tests for Metal *** Prices[11]

	Gold		Lead	
	No Jumps	**Jumps**	**No Jumps**	**Jumps**
OU				
μ	5.751 (0.096)	5.356 (0.983)	6.291 (0.102)	6.143 (0.131)
κ	0.008 (0.005)	0.002 (0.004)	0.013 (6.1E-3)	0.012 (5.3E-3)
σ	0.017 (5E-4)	0.012 (0.001)	0.033 (8.9E-4)	0.024 (1.3E-3)
λ		0.142 (0.045)		0.152 (0.048)
θ		0.003 (0.004)		0.011 (7.6E-3)
δ		0.032 (0.004)		0.055 (7.5E-3)
MLE	1807.30	1868.71	1371.61	1414.38
LR	122.81 [0.01, 0.01]		85.54 [0.01, 0.01]	
OU GARCH				
μ	5.453 (0.512)	5.375 (1.042)	6.292 (0.100)	5.918 (0.286)
κ	3.3E-3 (4.1E-3)	1.8E-3 (4.0E-4)	0.011 (5.6E-3)	8.8E-3 (5.5E-3)
ρ_0	2.1E-4 (1.6E-5)	1.4E-4 (1.7E-5)	4.4E-4 (1.0E-4)	1.9E-4 (6.6E-5)
ρ_1	0.335 (0.079)	0.068 (0.050)	0.201 (0.048)	0.165 (0.044)
ρ_2	0.000 (- -)	0.000 (- -)	0.346 (0.124)	0.411 (0.107)
λ		0.120 (0.050)		0.267 (0.101)
θ		0.000 (- -)		0.012 (6.1E-3)
δ		0.033 (5.4E-3)		0.035 (4.8E-3)
MLE	1828.45	1870.98	1416.58	1431.34
LR	85.05 [0.01, 0.01]		29.51 [0.01, 0.02]	

[11]Standard errors are in parenthesis. For the no-GARCH model, we report the bootstrap p-value $\widehat{p}_N(LR_0|\widehat{a}_1)$ [(2.10)] and the bounds p-value $\widetilde{p}_N(LR_0)$ [(2.15)]; for the GARCH-MRM, we report the bootstrap p-value $\widehat{p}_N(LR_0|\widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2)$ [(2.10)] $\widetilde{p}_N(LR_0)$ and the bound p-values $\widetilde{p}_N(LR_0)$ [(2.15)] respectively. Results with daily data for the Jump/MRM/GARCH model may be summarized as follows. Aluminium: estimated $\lambda = 0.089$ and $LR = 210.376$ (MC p-value = 0.01); Copper: estimated $\lambda = 0.128$ and $LR = 122.423$ (MC p-value = 0.01); Gold: estimated $\lambda = 0.330$ and $LR = 705.582$ (MC p-value = 0.01); Lead: estimated $\lambda = 0.097$ and $LR = 253.631$ (MC p-value 0.01). The MC p-values are based on (see (2.15)).

Table 4: Parameter Estimates and Bernoulli Jump Tests for Technology Stocks[12]

	Intel		Microsoft	
	No Jumps	**Jumps**	**No Jumps**	**Jumps**
GBM				
μ	1.1E-3 (4.7E-4)	1.7E-3 (4.7E-4)	1.3E-3 (4.0E-4)	1.4E-3 (3.9E-4)
σ	0.027 (3.3E-4)	0.021 (5.3E-4)	0.023 (2.8E-4)	0.019 (5.1E-4)
λ		0.122 (0.023)		0.106 (0.027)
θ		-4.8E-3 (3.7E-3)		-5.6E-4 (2.9E-3)
δ		0.049 (3.6E-3)		0.042 (3.8E-3)
MLE	7431.74	7608.31	7993.49	8138.51
LR	353.15 [0.01]		290.05 [0.01]	
GBM	**GARCH**			
μ	1.6E-3 (4.3E-4)	2.0E-3 (4.3E-4)	1.7E-3 (3.6E-4)	1.6E-3 (3.6E-4)
ρ_0	1.1E-5 (8.0E-6)	4.0E-6 (2.0E-6)	2.1E-5 (5.0E-6)	1.9E-5 (7.0E-6)
ρ_1	0.041 (0.016)	0.024 (6.3E-3)	0.073 (0.012)	0.062 (0.013)
ρ_2	0.945 (0.026)	0.963 (0.011)	0.889 (0.018)	0.877 (0.029)
λ		0.071 (0.019)		0.055 (0.018)
θ		-8.0E-3 (5.0E-3)		-2.1E-3 (4.4E-3)
δ		0.048 (5.1E-3)		0.045 (5.9E-3)
MLE	7609.68	7707.79	8136.69	8222.57
LR	196.20 [0.01, 0.01]		171.76 [0.01, 0.01]	

[12]Standard errors are in parenthesis. For the no-GARCH model, we report $\widehat{p}_N(LR_0)$ based on (2.8); for the GARCH-GBM, we report $[\widehat{p}_N(LR_0|\widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2), \widetilde{p}_N(LR_0)]$, i.e. the bootstrap and bounds p-values defined by (2.10) and (2.15) respectively. Results with weekly data for the Jump/GBM model (no-GARCH) may be summarized as follows. Intel: estimated $\lambda = 0.019$ and $LR = 36.974$ (MC p-value = 0.01); Microsoft: estimated $\lambda = 0.210$ and $LR = 52.163$ (MC p-value = 0.01); Oracle: estimated $\lambda = 0.028$ and $LR = 43.924$ (MC p-value = 0.01); Sun: estimated $\lambda = 0.062$ and $LR = 65.950$ (MC p-value 0.01). The MC p-values are based on (2.8).

Table 4 (Continued): Parameter Estimates and Bernoulli Jump Tests for Technology Stocks[13]

	Oracle		Sun	
	No Jumps	**Jumps**	**No Jumps**	**Jumps**
GBM				
μ	1.3E-4 (6.6E-4)	7.6E-4 (5.8E-4)	8.8E-4 (6.0E-4)	1.1E-3 (5.9E-4)
σ	0.038 (4.7E-4)	0.029 (6.2E-4)	0.035 (4.2E-4)	0.027 (7.4E-4)
λ		0.088 (0.015)		0.148 (0.027)
θ		5.8E-3 (4.9E-3)		-1.7E-3 (3.4E-3)
δ		0.084 (6.3E-3)		0.058 (4.1E-3)
MLE	6280.21	6575.99	6599.42	6764.54
LR	591.54 [0.01]		330.24 [0.01]	
GBM	**GARCH**			
μ	2.1E-3 (6.0E-4)	1.4E-3 (5.4E-4)	1.8E-3 (5.4E-4)	1.8E-3 (5.6E-4)
ρ_0	2.1E-5 (6.0E-6)	8.0E-6 (4.0E-6)	4.4E-5 (1.6E-5)	1.3E-5 (6.0E-6)
ρ_1	0.040 (8.4E-3)	0.017 (5.0E-3)	0.063 (0.016)	0.038 (0.008)
ρ_2	0.947 (0.011)	0.968 (0.011)	0.901 (0.028)	0.933 (0.016)
λ		0.045 (9.6E-3)		0.102 (0.026)
θ		1.7E-3 (6.4E-3)		-5.1E-3 (4.5E-3)
δ		0.100 (9.9E-3)		0.055 (5.6E-3)
MLE	6422.28	6683.29	6771.03	6868.42
LR	522.02 [0.01, 0.01]		194.78 [0.01, 0.01]	

---

[13]Standard errors are in parenthesis. For the no-GARCH model, we report $\widehat{p}_N(LR_0)$ based on (2.8); for the GARCH-GBM, we report $[\widehat{p}_N(LR_0|\widehat{a}_1, \widehat{\rho}_1, \widehat{\rho}_2), \widetilde{p}_N(LR_0)]$, i.e. the bootstrap and bounds p-values defined by (2.10) and (2.15) respectively. Results with weekly data for the Jump/GARCH/GBM model may be summarized as follows. Intel: estimated $\lambda = 0.003$ and $LR = 21.677$ (MC p-value = 0.01); Microsoft: estimated $\lambda = 0.069$ and $LR = 31.213$ (MC p-value = 0.01); Oracle: estimated $\lambda = 0.120$ and $LR = 52.699$ (MC p-value = 0.01); Sun: estimated $\lambda = 0.046$ and $LR = 27.876$ (MC p-value 0.01). The MC p-values is bound based (see (2.15)).

Appendix

A Fundamental definitions

This section provides the formal definition of an *exact* test (or p-value) and discusses some potential difficulties of implementing this definition.

Definition A.1 *Consider a test problem pertaining to a parametric model, i.e. the case where the data generating process [DGP] is determined up to a finite number of unknown real parameters $\omega \in \Omega$. Let Ω_0 refer to the subspace of Ω compatible with the null hypothesis H_0 under test. Without loss of generality, consider a test statistic with critical region $S \geq c$. To obtain an α-level test, c must be chosen so that*

$$\text{(A.3)} \qquad\qquad \sup_{\omega \in \Omega_0} P_\omega \left(S \geq c \right) \leq \alpha \ .$$

This test has size α if and only if

$$\text{(A.4)} \qquad\qquad \sup_{\omega \in \Omega_0} P_\omega \left(S \geq c \right) = \alpha \ .$$

Whereas size control, if possible, is desirable, level control is *required*. In the light of definition A.1, two difficulties must be dealt with to obtain an exact test. First, the null distribution of S must be derived, and S is typically a complicated function of the fundamentals. However, if simulated values of S which satisfy the null can be obtained [possibly conditional on nuisance parameters], procedures such as MC tests or the bootstrap can be used to tackle this first difficulty.

The second difficulty intervenes if the null distribution of the test statistic S used depends on parameters not set by the null hypothesis (nuisance parameters), so S is not pivotal. This happens when we test for jumps in jump-diffusion models, but also in many

other situations of practical interest. Indeed, provably pivotal statistics are rare beyond linear models. To deal with this difficulty, the MC test technique can be extended by maximizing the rejection probabilities over the relevant parameter space (Dufour 1995). For this - or any other size correction - approach to be useful, however, it is clear that the maximization problem should be bounded: the null distribution of the test statistic S should admit a nuisance parameter-free bound, which means that it should be *boundedly pivotal* (Dufour 1997). This property can be defined as follows.

Definition A.2 *Consider the test based on S underlying (A.4)-(A.3) and suppose that it is possible to find another pivotal statistic $\overline{S}$ such that*

$$(A.5) \qquad \qquad \forall \omega \in \Omega_0, \forall c, \ P_\omega(S \geq c) \leq P(\overline{S} \geq c)$$

under the null. Then S is said to be boundedly pivotal.

Combining (A.5) with (A.3), we see that a test based on S with $\overline{S}$'s critical points is level-correct (i.e. satisfy the level constraint (A.3)) given the pivotal character of $\overline{S}$. When the null distribution of $\overline{S}$ is non-standard yet simulatable, the MC test technique may be used to obtain the critical points of $\overline{S}$ by simulation (see Dufour and Khalaf 2001b, 2002a).

B   Monte Carlo tests

In this section, we present the fundamentals underlying the MC tests method [Dufour (2002)]. Consider the case where the null distribution of a statistic S at hand is simulatable, given a finite number of *known* real parameters. In other words, retaining the notational framework of the preceding section, let $\omega = \omega_0 \in \Omega$, where ω_0 is known.

1. Let S_0 denote the observed test statistic.

2. By Monte Carlo methods, obtain N *i.i.d.* draws form the null distribution of replications of S given ω; denote these simulated statistics S^j, $j = 1, \ldots, N$.

3. Compute the MC p-value

(B.6) $$\widehat{p}_N(S_0|\omega_0) = \frac{N\widehat{G}_N(S_0; S^j, \ j = 1, \ \ldots, \ N) + 1}{N + 1},$$

where $N\widehat{G}_N(S_0; S^j, \ j = 1, \ \ldots, \ N)$ is the number of simulated criteria $\geq S_0$ and

$$\widehat{R}_N(S_0) = N - N\widehat{G}_N(S_0; S^j, \ j = 1, \ \ldots, \ N) + 1$$

gives the rank of S_0 in the series S_0, S_1, ..., S_N. Formally, $\widehat{G}_N(x; S^j, \ j = 1, \ \ldots, \ N)$ obtains as defined in (??)-(??).

4. The MC critical region, conditional on ω_0 is

(B.7) $$\widehat{p}_N(S_0|\omega_0) \leq \alpha, \quad 0 < \alpha < 1.$$

(B.6) gives the *empirical probability* to observe a value as extreme or more extreme than S_0 under the null. Consequently, $\widehat{p}_N(S_0|\omega_0)$ may be viewed as a randomized MC p-value. Dufour (2002) proves that if $\alpha(N + 1)$ is an integer, $P_{(H_0)}[\widehat{p}_N(S_0) \leq \alpha] = \alpha$. This result rests on the following Lemma.

Lemma B.1 *Let* $Z_0, Z_1, \ldots, Z_N$ *be exchangeable real random variables, and let* R_j *be the rank of* Z_j *in the series* $\{Z_0, \ldots, Z_N\}$ *assuming a non-decreasing ordering. Then*

$$P[R_j/(N+1) \geq x] = \begin{cases} 1 & , \ if \ x \leq 0, \\ (1 + I[(N+1)(1-x)])/(N+1) & , \ if \ 0 < x \leq 1, \\ 0 & , \ if \ x > 1, \end{cases}$$

29

where $I(x)$ is the largest integer less than or equal to x.

If the intervening parameter ω is unknown, MC tests are generally based on the critical region

(B.8)
$$\sup_{\omega \in \Omega_0} \left[\widehat{p}_N(S_0|\omega) \right] \leq \alpha.$$

Specifically, Dufour (2002) demonstrates that the test (henceforth denoted *maximized Monte Carlo* (MMC) test) based on the latter critical region is exact at level α, in the sense that

(B.9)
$$\left\{ P_{(H_0)} \left[\widehat{p}_N(S_0|\omega) \right] \right\} \leq \alpha.$$

Finally, observe that since no asymptotics on N was used to derive all the above, the number of MC replications needs not be very large. In fact a value as small as 19 is necessary to control size; of course, power increases with N, yet in the literature on MC tests (see the references above), it is demonstrated that 99 replications is a quite reasonable choice.

C   Likelihood functions

This section provides the expressions of the log-likelihood functions for the mean-reverting models used in the empirical illustration. $\phi(z)$ designates the density of the standard normal distribution.

We suppose here that $X = \ln(P)$ follows the Ornstein-Uhlenbeck process: $dX = \kappa(\mu - X)dt + \sigma dw$, where dw is an increment of a standardized Wiener Process. From Karlin and Taylor (1981), we know that, conditional on $X = x_0$ at time 0, X_T is normally distributed with mean $\mu + (x_0 - \mu)e^{-\kappa T}$ and variance $\sigma^2 \frac{(1 - e^{-2\kappa T})}{2\kappa}$. If at time t $(0 < t < T)$

we add a normally distributed jump (with mean θ and variance σ^2) to X_t, then X_T (conditional on the arrival of a $N(\theta, \sigma^2)$ jump at t) is normally distributed with mean $\mu + (x_0 - \mu)e^{-\kappa T} + \theta e^{-\kappa(T-t)}$ and variance $\sigma^2 \frac{(1-e^{-2\kappa T})}{2\kappa} + \delta^2 e^{-2\kappa(T-t)}$. To derive $f_{X_T}(\cdot)$, the unconditional (with respect to the time of arrival of a jump) distribution of X_T for a single jump occurring between 0 and T, we assume that the density of the arrival of jumps is the rescaled exponential distribution $\frac{\lambda e^{-\lambda t}}{1-e^{-\lambda T}}$. We obtain:

$$f_{X_T}(x) = \int_0^T \frac{\phi\left(\frac{x_t - \mu(1-e^{-\kappa T}) - x_{t-1}e^{-\kappa T} - \theta e^{-\kappa(T-t)}}{\sqrt{\sigma^2 \frac{1-e^{-2\kappa T}}{2\kappa} + \delta^2 e^{-2\kappa(T-t)}}}\right)}{\sqrt{\sigma^2 \frac{1-e^{-2\kappa T}}{2\kappa} + \delta^2 e^{-2\kappa(T-t)}}} \frac{\lambda e^{-\lambda t}}{1 - e^{-\lambda}} dt \Bigg]$$

To simplify this expression, we assume that, if there is a jump in the interval $(0, 1)$ it takes place at time 1. Since the Monte-Carlo method requires many evaluations of the likelihood function, this assumption reduces computing time substantially. Thus, the log-likelihood function for an Ornstein-Uhlenbeck process with Bernoulli jumps arriving with frequency λ is given by:

$$l_B(\overrightarrow{\Theta_B}; \vec{x}) = \sum_{t=1}^T ln\left[(1-\lambda)\frac{\phi\left(\frac{x_t - h_0 - h_1 x_{t-1}}{h_2}\right)}{h_2} + \lambda\frac{\phi\left(\frac{x_t - h_0 - h_1 x_{t-1} - \theta}{\sqrt{h_2^2 + \delta^2}}\right)}{\sqrt{h_2^2 + \delta^2}}\right]$$

$$\overrightarrow{\Theta_B} = (\mu, \kappa, \sigma^2, \lambda, \theta, \delta^2), \ \kappa \geq 0, \ \lambda \geq 0,$$

$$h_0 = \mu(1 - e^{-\kappa}), \ h_1 = e^{-\kappa}, \ h_2 = \sigma\left(\sqrt{\frac{1 - e^{-2\kappa}}{2\kappa}}\right), \ h_3 = \theta e^{-\kappa}, \ h_4 = \delta^2 e^{-2\kappa}.$$

Adding a GARCH(1,1) error structure leads to the log-likelihood function for an

Ornstein-Uhlenbeck with Bernoulli jumps and GARCH(1,1) errors:

$$l_{BG}(\overrightarrow{\Theta_{BG}}; \vec{x}) = \sum_{t=1}^{T} ln\left[(1-\lambda)\frac{\phi\left(\frac{x_t - h_0 - h_1 x_{t-1}}{\sqrt{h_t}}\right)}{\sqrt{h_t}} + \lambda\frac{\phi\left(\frac{x_t - h_0 - h_1 x_{t-1} - \theta}{\sqrt{h_t + \delta^2}}\right)}{\sqrt{h_t + \delta^2}}\right],$$

$$
\begin{aligned}
\overrightarrow{\Theta_{BG}} &= (\mu, \kappa, \rho_0, \rho_1, \rho_2, \lambda, \theta, \delta^2),\ \kappa \geq 0,\ \rho_0 > 0,\ \rho_1 \geq 0,\ \rho_2 \geq 0,\ \lambda \geq 0, \\
h_0 &= \mu(1 - e^{-\kappa}),\ h_1 = e^{-\kappa},\ h_3 = \theta e^{-\kappa},\ h_4 = \delta^2 e^{-2\kappa}, \\
h_t &= \rho_0 + \rho_1(x_{t-1} - h_0 - h_1 x_{t-2})^2 + \rho_2 h_{t-1},\ t = 3, ..., T.
\end{aligned}
$$

For our empirical illustration, we find that our assumption on the time of arrival of jumps has no significant impact on the results.

References

[1] Andrews, D. W. K. (2000): "Inconsistency of the Bootstrap When a Parameter is on the Boundary of the Parameter Space," *Econometrica* 68, 399-405.

[2] Andrews, D. W. K. (2001): "Testing When a Parameter is on the Boundary of the Maintained Hypothesis," *Econometrica* 69, 683-734.

[3] Ahn, C. M. and Thompson, H. E. (1988): "Jump-Diffusion Processes and the Term Structure of Interest Rates," *Journal of Finance* 43, 155-174.

[4] Akgiray V. and Booth, G. G. (1988): "Mixed Diffusion-Jump Process Modelling of Exchange Rate Movements," *The Review of Economics and Statistics 70, 631-637.*

[5] Bakshi, G., Cao C. and Z. Chen (1997). "Empirical Performance of Alternative Option Pricing Models," *The Journal of Finance* 52, 2003-2049.

[6] Bakshi, G., Cao C. and Z. Chen (2000). "Pricing and Hedging Long-Term Options Models," *Journal of Econometrics* 94, 277-318.

[7] Ball, C. and Torous, W. (1985): "On Jumps in Common Stock Prices and their Impact on Call Option Pricing," *Journal of Finance* 40, 155-173.

[8] Bates, D. S. (1996a): "Dollar Jump Fears, 1984-1992: Distributional Abnormalities Implicit in Currency Futures Options," *Journal of International Money and Finance* 15, 65-93.

[9] Bates, D. S. (1996b): "Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options," *The Review of Financial Studies* 9, 69-107.

[10] Bates, D. S. (2000): "Post-87 Crash Fears in S&P 500 Futures Options", *Journal of Econometrics* 94, 181-238.

[11] Beck S, (2001). "Autoregressive Conditional Heteroskedasticity in Commodity Spot Prices," *Journal of Applied Econometrics* 16, 115-132.

[12] Benzoni L, (2001). "Pricing Options under Stochastic Volatility: an Empirical Investigation," Working Paper, Finance Department, Carlson School of Management, University of Minnesota.

[13] Bera, A. K. and Ra, S. (1995): "A Test for the Presence of Conditional Heteroskedasticity within ARCH-M Framework," *Econometric Reviews* 14, 473-485.

[14] Brorsen, B. W. and Yang, S. R. (1994): "Nonlinear Dynamics and the Distribution of Daily Stock Index Returns," *The Journal of Financial Research* 17, 187-203.

[15] Chernov, M., Gallant, A. R., Ghysels E. and G. Tauchen (2002): Alternative Models for Stock price Dynamics, working paper, Columbia University.

[16] Campbell, J. W., Lo A. W. and MacKinlay A. C. (1997): *The Econometrics of Financial Markets,* Princeton University Press, Princeton, New Jersey.

[17] Dagenais, M. G. and Dufour, J.-M. (1991): "Invariance, Nonlinear Models, and Asymptotic Tests," *Econometrica* 59, 1601-1615.

[18] Das S. R. (2002): "The surprise element: jumps in interest rates," *Journal of Econometrics*, 106, 27-65.

[19] Drost, F. C., and Nijman, T. E. (1993): "Temporal Aggregation of GARCH Processes," *Econometrica* 61, 909-927.

[20] Drost, F. C., Nijman, T. E. and Werker, B. J. M. (1998): "Estimation and Testing in Models Containing Both Jump and Conditional Heteroskedasticity," *Journal of Business and Economic Statistics* 16, 237-243.

[21] Duan, J.-C. (1997): "Augmented GARCH(p,q) Process and its Diffusion Limit," *Journal of Econometrics* 79, 97-127.

[22] Dufour, J.-M. (1989). "Nonlinear hypotheses, inequality restrictions and non-nested hypotheses: exact simultaneous tests in linear regressions", *Econometrica* 57, 335-355.

[23] Dufour, J.-M. (1997): "Some Impossibility Theorems in Econometrics with Applications to Structural and Dynamic Models," *Econometrica* 65, 1365-1387.

[24] Dufour, J.-M., Farhat, A., Gardiol, L., and Khalaf, L. (1998): "Simulation-Based Finite Sample Normality Tests in Linear Regressions," *The Econometrics Journal* 1, 154-173.

[25] Dufour, J.-M. (2002): "Monte Carlo Tests in the Presence of Nuisance Parameters with Econometric Applications," *Journal of Econometrics*, forthcoming.

[26] Dufour, J.-M., and Khalaf, L. (2001a): "Monte Carlo Test Methods in Econometrics," in Baltagi B. (Ed.) *Companion to Theoretical Econometrics*, Blackwell; 494-519.

[27] Dufour, J.-M., Khalaf, L., Bernard J.-T. and I. Genest (2001b). "Simulation-based finite-sample tests for heteroskedasticity and ARCH effects", Working Paper, Centre de recherche et développement en économique (C.R.D.E.) and Département de sciences économiques, Université de Montréal.

[28] Dufour, J.-M., and Khalaf, L. (2002a): "Exact Tests for Contemporaneous Correlation of Disturbances in Seemingly Unrelated Regressions", *Journal of Econometrics*, 106, 143-170.

[29] Dufour, J.-M., and Khalaf, L. (2002b): "Simulation-Based Finite and Large Sample Tests in Multivariate Regressions", *Journal of Econometrics*, forthcoming.

[30] Hansen, B. E. (1996): "Inference When a Nuisance Parameter is not Identified Under the Null Hypothesis," *Econometrica* 64, 413-430.

[31] Jarrow, R. A. and Rosenfeld, E. R. (1984): "Jump Risks and the Intertemporal Capital Asset Pricing Model," *The Journal of Business* 57, 337-351.

[32] Jorion, P. (1988): "On Jump Processes in the Foreign Exchange and Stock Markets," *The Review of Financial Studies* 1, 427-445.

[33] Karlin, S. and H. M. Taylor (1981): *A Second Course in Stochastic Processes*, Academic Press, San Diego.

[34] Merton, R. C. (1976): "Option Pricing When Underlying Stock Returns are Discontinuous," *Journal of Financial Economics* 3, 125-144.

[35] Merton, R. C. (1990): *Continuous Time Finance*, Blackwell, Cambridge Massachusetts.

[36] Pan J. (2002): "The Jump-Risk Premia Implicit in Options: Evidence from an Integrated Time-Series Study," *Journal of Financial Economics* 63, 3-50.

[37] Saphores, J.-D, Khalaf, L. and D. Pelletier (2002). "On Jumps and ARCH Effects in Natural Resource Prices. An Application to Stumpage Prices from Pacific Northwest National Forests ", *American Journal of Agriculture Economics*, 84(2), 387-400.

[38] Schwartz, E. S. (1997): "The Stochastic Behavior of Commodity Prices: Implications for Valuation and Hedging," *Journal of Finance* 52, 923-973.

[39] Staiger, D. and J.H. Stock (1997). "Instrumental Variables regression with weak instruments", *Econometrica* 65, 557-586.

[40] Vlaar, P. J. G., and F. C. Palm (1993): "The Message in Weekly Exchange Rates in the European Monetary System: Mean Reversion, Conditional Heteroskedasticity, and Jumps," *Journal of Business and Economic Statistics* 11 (3), 351-360.

[41] Wang J., and E. Zivot (1998). "Inference on Structural Parameters in Instrumental Variable Regression with Weakly Correlated Instruments", *Econometrica* 66, 1389-1404.