Testing for Parameter Instability in Competing Modeling Frameworks*

Francesco Calvori[a], Drew Creal[b], Siem Jan Koopman[c], André Lucas[c]

[(a)] *Department of Statistics "G. Parenti", University of Florence*

[(b)] *Booth School of Business, University of Chicago*

[(c)] *VU University Amsterdam and Tinbergen Institute*

Abstract

We develop a new parameter stability test against the alternative of observation driven generalized autoregressive score dynamics. The new test generalizes the ARCH-LM test of Engle (1982) to settings beyond time-varying volatility and exploits any autocorrelation in the likelihood scores under the alternative. We compare the test's performance with that of alternative tests developed for competing time-varying parameter frameworks, such as structural breaks and observation driven parameter dynamics. The new test has higher and more stable power against alternatives with frequent regime switches or with non-local parameter driven time-variation. For parameter driven time variation close to the null or for infrequent structural changes, the test of Müller and Petalas (2010) performs best overall. We apply all tests empirically to a panel of losses given default over the period 1982–2010 and find significant evidence of parameter variation in the underlying beta distribution.

Key words: time-varying parameters; observation driven models; parameter driven models; structural breaks; generalized autoregressive score model; regime switching; credit risk.

JEL classifications: C12, C52, C22.

*We have benefitted from the comments of participants of the 2013 Workshop on Score Driven Models, Amsterdam. Drew Creal thanks the William Ladany Faculty Scholar Fund at the Booth School of Business for financial support. André Lucas thanks the Dutch National Science Foundation (NWO), vici grant 453-09-005 for financial support. Email addresses: calvori@disia.unifi.it, drew.creal@chicagobooth.edu, s.j.koopman@vu.nl, a.lucas@vu.nl.

1 Introduction

A key concern in empirical model building is model stability. Hansen (2001) provides an overview of a large number of different parameter stability tests found in the literature, including standard tests such as the Chow (1960) break test, the supremum F-tests of Andrews (1993), and the weighted F-tests by Andrews and Ploberger (1994).

When testing for parameter stability, the model under the alternative hypothesis of parameter instability can take many different forms. For example, there might be one or more deterministic structural breaks in the parameters of a model as in, for example, Vogelsang and Perron (1998), Bai and Perron (2003), Perron (2006), and Qu and Perron (2007); the parameters might exhibit regular regime switches as in Hamilton (1989); or the parameters might evolve continuously over time, either in a parameter driven (state space) framework such as Harvey (1989), Bauwens and Veredas (2004), Shephard (2005), and Hafner and Manner (2012), Durbin and Koopman (2012), or an observation driven framework such as Engle (1982), Bollerslev (1986), Engle and Russell (1998), Davis et al. (2003), Patton (2006), and Creal et al. (2013).

The goal of this paper is twofold. First, we develop a new test for parameter stability in non-linear, non-Gaussian models against the generalized autoregressive score (GAS) alternative of Creal et al. (2013). GAS models are a flexible class of observation driven time-varying parameter models characterized by a parametric conditional observation density. As their likelihood is available in closed form, likelihood based estimation and inference is straightforward. The usefulness of the GAS framework to capture time-variation in parameters has been illustrated in for example Creal et al. (2011) and Harvey (2013) for multivariate volatility and correlation models, in Creal et al. (2014) for mixed measurement factor models, and in Oh and Patton (2013) and De Lira Salvatierra and Patton (2013) for copula models. GAS models generalize other well known observation driven time varying parameter models such as the GARCH model of Engle (1982) and Bollerslev (1986) or the ACD model of Engle and Russell (1998). Moreover, their out-of-sample forecasting accuracy is at par with that of comparable non-linear non-Gaussian state space models; see Koopman et al. (2012). A parameter stability test against the alternative of GAS dynamics can thus provide a useful signal that a static model is too simplistic and needs to be augmented.

The Lagrange multiplier (LM) test we develop takes a highly intuitive form: it tests for non-zero autocorrelations in the likelhood score of the static model. As such, it can be seen as an omnibus diagnostic tool generalizing the familiar ARCH-LM test of Engle (1982) to settings beyond time-varying volatility. The asymptotic distribution of the test easily follows by familiar

results from White (1987). Moreover, similar to most omnibus LM diagnostic tests, the new test can easily be computed by means of an auxiliary regression.

As a second goal of our paper, we investigate the finite sample properties of the new test in relation to that of other familiar (and less familiar) tests from the literature. We consider linear as well as non-linear models. Computational efficiency then becomes a concern, as estimating an additional non-linear model is typically costly. This automatically favors the use of LM-based test statistics. In particular, we consider the sup-LM tests of Andrews (1993), the test against random walk parameter alternatives of Nyblom (1989), and the recent test against local parameter driven time variation of Müller and Petalas (2010). Each of these tests is built with an entirely different alternative time-varying parameter framework in mind than the GAS-based test. All tests are applicable to both linear and non-linear settings.

The results of Müller and Petalas (2010) are particularly interesting for our paper. They prove that a test against a parameter driven alternative is asymptotically locally optimal against an alternative where the magnitude of the time variation in the unstable parameters shrinks as the sample size grows. Their theoretical results are supported by numerical simulations; see also Elliott and Müller (2006) for the case of linear models. We extend their results in several directions. First, we consider a wider set of alternatives for parameter variation, including deterministic regime switches, random walk parameters, and stationary parameter driven dynamics. Second, we pay attention to non-local alternatives. The theoretical results in Müller and Petalas (2010) do not make clear predictions about the behavior of alternative testing methodologies under non-local alternatives in finite samples. In particular, tests against other than parameter driven alternatives might have a better overall performance in a finite sample setting. In addition, the finite sample performance of the different testing procedures may crucially depend on the type of data generating process under the alternative.

Our simulation experiment provides several interesting results. The new test against the observation driven (GAS) alternative exhibits higher power for alternatives that display regular regime switches or non-local parameter driven time-variation. For parameter driven time variation close to the null or for infrequent structural changes, the test of Müller and Petalas (2010) performs best. For local time variation, the optimality of the Müller and Petalas (2010) approach follows directly from their analytical results. Also the good performance in the setting with infrequent breaks is in line with their simulation results, where they show how their test typically performs similar to or better than the sup-LM test of Andrews (1993). If the size of the time-variation is larger or more frequent, however, the new GAS-LM test against observation driven alternatives performs better. The new test's power performance is also surprisingly robust over alternative specifications of the data generating process, in contrast to most of its

competitors.

The explanation of our findings follows from the construction of the different tests. The GAS-LM test builds directly on the autocorrelations of the likelihood scores in the static model. These autocorrelations can typically be estimated rather well even if the time varying parameter in the unknown data generating process moves quickly. This explains the good performance of the test in settings with many breaks or strong mean reverting parameter dynamics. The test of Müller and Petalas (2010), by contrast, uses the unconditional volatility of the path of the time varying parameter as its main ingredient. This path is estimated under a high persistence assumption. If the true time varying parameter moves quickly, the estimated path typically becomes almost constant, thus reducing the estimated volatility of the path and consequently the power of the test. The same holds for the path estimated under GAS dynamics, but this does not affect the GAS-LM test as the latter is not based on the volatility of the estimated path of the time varying parameter. For slow time variation, the opposite holds and the likelihood ratio (LR) perspective of the Müller and Petalas (2010) test results in a better performance of the test compared to the LM based GAS test.

We apply all tests empirically to test for the stability of the parameters of loss given default (LGD) models in a credit risk context. LGD is the fraction of the outstanding amount of a loan or bond that is lost in case the company gets into default. The LGD is a key ingredient of current financial risk management and regulation. Many financial industry credit risk models for LGDs use static parameters. A prime example is the use of a static beta disribution for LGDs. Such a modeling strategy is highly risky if the properties of LGDs actually vary over time. For example, losses could be on average higher in situations where default risk is also higher, thus exacerbating total expected losses defined as the probability of default times the LGD. If such time-variation is a property of the data, it should be modeled and would typically result in higher capital requirements for financial instututions. We use a panel data set of LGDs for corporate bond data obtained from Moody's to test for the presence of such time-variation. The data set is non-standard and therefore provides an interesting example of the flexibility of our testing approach. Using quarterly data, the number of LGD observations per quarter varies over time. This follows directly from the fact that the LGD can only be observed in case of a default, where the number of defaults varies over time. Assuming the LGDs are drawn from a beta distribution with possibly time-varying parameters, all our tests strongly confirm that the distributional properties of LGDs vary over time. In particular, we find that LGDs were on average very low compared to the static model during the period leading up to the 2008 financial crisis, suggesting that the abundance of liquidity during this period not only prevented firms from defaulting, but also mitigated the losses for those cases in which a default

was unavoidable.

The remainder of this paper is set up as follows. In Section 2 we describe our new test statistic as well as the main alternative tests from the literature. In Section 3 we describe our Monte-Carlo experiment to study the finite sample properties of the tests and present the simulation results. In Section 4, we apply the different tests to our empirical panel of corporate bond loss fractions. Section 5 concludes. An online appendix to this paper holds additional simulation results.

2 Testing frameworks for time-varying parameters

In this section we present our three different frameworks to test for parameter instability. Each of these frameworks has been designed with a particular alternative in mind for the dynamic behavior of the time-varying parameters. As we focus on tests that should be applicable to non-linear models, we focus on Lagrange Multiplier (LM) type tests only. Although LM tests are typically less powerful against specific alternatives than Wald or likelihood ratio type tests, LM tests only require estimation under the null. In a non-linear, non-Gaussian setting, this provides considerable advantages. Estimating non-linear models under the alternative can prove cumbersome if not infeasible, for example, if one considers multiple break dates or if the likelihood under the alternative is not known in closed form.

To fix the notation, we consider a dependent variable $y_t \in \mathbb{R}^m$ for $t = 1, \ldots, T$, where T denotes the sample size; a vector of time-varying parameters $f_t \in F \subset \mathbb{R}^k$, and a vector of static parameters $\delta \in D \subset \mathbb{R}^n$, where F and D denote the parameter space of the time-varying and static parameter vectors, respectively.

2.1 Observation driven time-variation

In an observation driven framework, the time-varying parameter f_t is driven by a deterministic function of lagged dependent variables and contemporaneous or lagged exogenous variables. The observation driven modeling framework has the advantage that the likelihood is available in closed form and can easily be evaluated through a prediction error decomposition. This leads to simple estimation and inference procedures. The main challenge in the observation driven framework is to determine which function of the observations to choose as a driver for f_t. A general approach encompassing many popular non-linear, non-Gaussian dynamic models is the generalized autoregressive score (GAS) model of Creal et al. (2013). GAS models use the (scaled) score of the conditional observation density to drive the parameter f_t through time. At each time step, the dynamics of the time-varying parameter can be interpreted as a steepest-

ascent or Gauss-Newton improvement, where we improve the local fit of the model by taking into account the information in the most recent observation and its distribution. The GAS framework encompasses the Gaussian GARCH model of Engle (1982) and Bollerslev (1986), the ACD and ACI models of Engle and Russell (1998) and Russell (2001), the MEM model of Engle and Gallo (2006) and Cipollini et al. (2012), models for Poisson counts in Davis et al. (2003), and the Beta-t-GARCH model of Harvey (2013), among many others.

In the GAS(p, q) framework, the time-varying parameter follows the dynamic specification

$$f_{t+1} = (\mathrm{I} - B_1 - \ldots - B_p)\omega + \sum_{i=1}^{q} A_i s_{t-i+1} + \sum_{j=1}^{p} B_j f_{t-j+1}, \tag{1}$$

where the elements of the vector ω and of the matrices A_i and B_j are static parameters for $i = 1, \ldots, p$ and $j = 1, \ldots, q$,

$$s_t := S_t \cdot \nabla_{f,t} := S_t \cdot \frac{\partial \ln p(y_t|f_t;\ \delta)}{\partial f_t}, \tag{2}$$

with $p(y_t|f_t;\ \delta)$ the conditional observation density, and $\nabla_{f,t}$ its score. The $k \times k$ matrix $S_t = S(f_t; \delta)$ scales the score by using, for example, a power of the Fisher information matrix of the conditional observation density to account for the curvature of the score; see Creal et al. (2013) for more details.

To develop the GAS-LM test, we draw the analogue with the ARCH-LM test of Engle (1982) or the GARCH-LM test of Lee (1991). The ARCH(1)-LM test of Engle for the model $y_t = x_t'\beta + \sigma_t \varepsilon_t$ with $\varepsilon_t \sim (0, 1)$ tests the null of a constant variance against the alternative

$$\sigma_{t+1}^2 = (1 - A)\omega + A\varepsilon_t^2 = (1 - A)\omega + A(\varepsilon_t^2 - \sigma_t^2) + A\sigma_t^2, \tag{3}$$

such that under the alternative σ_{t+1}^2 varies around the static level ω as driven by the scaled score of a Gaussian density, $\varepsilon_t^2 - \sigma_t^2$. Under the null hypothesis, $A = 0$.

The LM test against a GAS alternative takes the same perspective as (3), but acknowledges the fact that the time varying parameter f_t may characterize a different distributional property than the variance, and moreover that under the alternative the dynamics of f_t are driven by the score s_t of the conditional observation density rather than by the score $(\varepsilon_t^2 - \sigma_t^2)$ of the Gaussian volatility model in (3). Similar to (3), we thus test the null of no parameter variation against the GAS alternative

$$f_{t+1} = (\mathrm{I} - A_1 - \ldots - A_q)\omega + \sum_{i=1}^{q} A_i s_{t-i+1} + \sum_{i=1}^{q} A_i f_{t-i+1}. \tag{4}$$

Similar to (3), the time varying parameter f_t varies under the alternative around its static level ω. We can use the same arguments as in Lee (1991) to allow for different coefficients B_i (rather than A_i) for the lags of f_{t-i+1} under the alternative.

To define the LM test statistic, let $\ell_t(\delta, \omega, a) = \ln p(y_t|f_t;\ \delta)$ be the likelihood at time t, where we suppressed the dependence of f_t on the static parameters δ, ω, and $a = \text{vec}(A_1, \ldots, A_q)$. Define $\bar{s}_{p,t} = \text{vec}(s_t, \ldots, s_{t-p+1}))$, where $\iota_q \in \mathbb{R}^{q\times 1}$ is a vector of ones, and $\otimes$ is the Kronecker product. Also let $\mathcal{G}_t' = (\nabla_{\delta,t}', \nabla_{\omega,t}', \nabla_{\omega,t}' \otimes \bar{s}_{p,t-1}')$, with $\nabla_{\delta,t}$ and $\nabla_{\omega,t}$ denoting the derivatives of ℓ_t with respect to δ and ω, respectively. Following White (1987), the LM test for $\text{H}_0 : a = 0$ versus the alternative $\text{H}_1 : a \neq 0$, is given by

$$LM = \mathcal{G}'\mathcal{H}^{-1}\mathcal{G}, \quad \mathcal{G} = \sum_{t=1}^{T} \mathcal{G}_t, \quad \mathcal{H} = \sum_{t=1}^{T} \mathcal{G}_t\mathcal{G}_t', \tag{5}$$

where all derivatives are evaluated at the maximum likelihood estimates under the null hypothesis. As always, the covariance matrix $\mathcal{H}$ can be replaced by a robust long-term covariance matrix if needed, i.e.,

$$\tilde{\mathcal{H}} = \sum_{t=1}^{T}\sum_{\tau=1}^{t} w_{T,t-\tau}\left(\mathcal{G}_t\mathcal{G}_\tau' + \mathcal{G}_\tau\mathcal{G}_t'\right),$$

for some kernel weights $w_{T,t-\tau}$, see Andrews (1991).

Under the null and under standard regularity conditions, the GAS-LM test converges to a χ^2 distributed random variable with $\dim(a)$ degrees of freedom; see White (1987). The asymptotic statistical theory of the GAS-LM test is therefore entirely standard, in contrast to that of some alternative parameter stability tests.

Following Davidson and MacKinnon (1990), we can also write the LM test statistic as the explained sum of squares of the auxiliary OLS regression

$$1 = (\nabla_{\delta,t}', \nabla_{\omega,t}', \nabla_{\omega,t}' \otimes \bar{s}_{p,t-1})\beta_{LM} + \text{residual}, \tag{6}$$

where β_{LM} is a vector of auxiliary regression parameters, and all derivatives are again evaluated under the null. The regression interpretation of the GAS test makes it easy to compute in standard packages. The only quantities needed are the first derivatives of the conditional observation density at each time t, which we can easily obtain either analytically or numerically.

The GAS-LM test has an intuitive interpretation. Looking at the auxiliary regression (6), we see that the key term in the regression is $\nabla_{\omega,t} \otimes \bar{s}_{p,t-1}$. The elements of this vector are $\text{vec}(S_{t-i}\nabla_{\omega,t-i}\nabla_{\omega,t}')$ for $i = 1, \ldots, q$, where we have used the fact that under the null the score of the conditional density with respect to f_t is the same as that with respect to ω. The LM test against the GAS alternative thus checks whether there is any autocorrelation in the scores $\nabla_{f,t}$ of the *static* model. Such autocorrelations can be exploited to improve the fit of the model by using the likelihood scores as drivers for the time varying parameter as is done in the GAS framework.

Even though the above LM test has been derived with the GAS alternative in mind, we expect this test to also have power against alternative forms of parameter instability. As such

it can be regarded as an omnibus test for model misspecification. The same holds for the tests against structural breaks and against parameter driven time variation, which we discuss next.

2.2 Parameter driven time-variation

In parameter driven time-varying parameter models, the parameter f_t is a stochastic process that is subject to its own source of error. Important examples of this class of models are linear, Gaussian state space models (see, e.g. Harvey (1989)), stochastic volatility models as reviewed in Shephard (2005), stochastic conditional duration models as in Bauwens and Veredas (2004), and stochastic copula models such as Hafner and Manner (2012). The additional randomness in f_t on top of the randomness in y_t itself (conditional on f_t) makes these models harder to estimate. The likelihood function is typically not available in closed form except in specific cases, such as linear-Gaussian state space models and discrete-state hidden Markov models, see Durbin and Koopman (2012) and Hamilton (1989). Estimation of these models by likelihood methods typically requires approximation and/or simulation techniques; see for example Creal (2012) and Durbin and Koopman (2012) for a discussion of alternative approaches.

Müller and Petalas (2010), denoted as MP10 from now on, provide an elegant and generic set-up to test for parameter instability. Their approach encompasses non-linear and non-Gaussian models with moderately time-varying parameters. If the time variation vanishes asymptotically, MP10 show that we can address the inference problem about parameter stability by considering a linear Gaussian state space model where the observations are replaced by the likelihood scores of the static model. Moreover, they prove that such an approach is not only asymptotically optimal against the alternative of (local) parameter driven time varying parameters, but also against a much wider range of alternative (local) parameter dynamics. As such, the test stands in a long tradition of point optimal tests against local alternatives, such as random walk parameters; see for example Nyblom and Mäkeläinen (1983), Franzini and Harvey (1983), King and Hillier (1985), Nyblom (1989), and Elliott and Müller (2006).

The key intuition for the MP10 test follows from a pseudo-model

$$\begin{aligned} \mathcal{H}\mathcal{V}^{-1}\nabla_{\omega,t} &= \mathcal{S}^{-1}(f_t - \bar{f}) + \nu_t, & \nu_t &\sim \mathrm{N}(0, \mathcal{S}^{-1}) \\ (f_{t+1} - \bar{f}) &= (1 - cT^{-1})(f_t - \bar{f}) + \tilde{\nu}_t, & \tilde{\nu}_t &\sim \mathrm{N}(0, c^2T^{-2}\mathcal{S}^{-1}), \end{aligned} \tag{7}$$

where $\bar{f}$ is a fixed benchmark level for f_t around which there is local time variation, c_i is a fixed tuning parameter, $\mathcal{H} = T^{-1}\sum_{t=1}^{T}\partial^2 \ln p(y_t|\bar{f})/\partial f_t^2$, $\mathcal{V} = T^{-1}\sum_{t=1}^{T}\nabla_{\omega,t}\nabla'_{\omega,t}$, and $\mathcal{S} = \mathcal{H}^{-1}\mathcal{V}\mathcal{H}^{-1}$. The local time variation in f_t clearly vanishes as the sample size grows. Equation (7) can be seen as a linear Gaussian state space model, where the observations are the likelihood scores and the state $(f_t - \bar{f})$ is a nearly-integrated process.

The pseudo-observation model in (7) is close to the Laplace transformations used in the approximating linear state space models of Durbin and Koopman (2000) and Richard and Zhang (2007). In these papers, the approximating models are used to estimate the parameters in non-linear and non-Gaussian state space models via importance sampling. As explained in Müller and Petalas (2010), the primary difference between the approximating linear Gaussian model in these papers and the approximating model in MP10 is the use of the global Hessian H rather than the local Hessian of the conditional observation density at time t in the MP10 approach.

MP10 construct a point optimal test using a likelihood ratio test of the null $c = 0$ versus the alternative $c = 10$. Though the theory in MP10 is highly advanced, the proposed test statistic is actually surprisingly straightforward to compute using simple regression techniques. An algorithm is provided in the paper. The point optimality of the test gives it a clear likelihood ratio interpretation. As a result, the test has a clear power advantage compared to an LM based test. This stems from the fact that we actually obtain an approximate fit of the model under the (local) parameter driven alternative $c = 10$ based on the regressions used to compute the test statistic. This allows us to capture part of the corresponding gain of the likelihood ratio compared to the Lagrange multiplier test, just as if we would have estimated the model under both the null and the alternative rather than under the null only.

Based on similar regressions as the ones used to obtain the test statistic itself, MP10 also propose an algorithm to estimate the path of the time-varying parameter f_t. Their final estimate is a weighted average risk based combination of the estimated paths for different values of c, namely $c = 0, 5, 10, \ldots, 50$. We will use this estimate in our empirical application later on.

The crucial ingredient in the test of Müller and Petalas (2010) is the sum of $(f_t - \bar{f}) \cdot \nabla_{\omega,t}$. The test thus makes direct use of the variability of the estimated path f_t for $c = 10$ in deviation from its static counterpart $\bar{f}$. The differences $f_t - \bar{f})$ are weighted by the likelihood score with respect to the (possibly) dynamic parameter. If the estimated path f_t is relatively constant, or if the likelihood is not very sensitive with respect to f_t, the resulting test statistic is small. Note that the (smoothed) estimate of the path for the test is obtained under $c = 10$, which typically implies a high degree of persistence. This can become problematic if there is rapid time variation in f_t under the alternative, such as in the case of regular regime switches or strongly mean reverting parameter dynamics. In those cases, the estimated path of f_t can turn out rather constant, thus resulting in a low value of the test statistic and a low power of the corresponding testing procedure.

Compared to the GAS-LM test from Section 2.1, the MP10 test has three main differences. First, due to the choice of $c = 10$ and the structure of the auxiliary regressions, the MP10

test statistic weights both present and future autocovariances of the score. By contrast, the GAS-LM test only uses past autocovariances. Second, the GAS-LM test allows the user to include an explicit number of autocovariances through the choice of the parameter q. The MP10 test, by contrast, takes all autocovariances into account, but implicitly defines their weight through the choice of the parameter $c = 10$. Third, the distributions of the GAS and MP10 test under the null differ markedly. Whereas the GAS-LM test follows the standard χ^2 asymptotics of White (1987), the MP10 test follows the asymptotic distribution as derived in Elliott and Müller (2006).

2.3 Structural breaks

Andrews (1993) proposes a general parameter stability test for nonlinear parametric models against alternatives with a one-time structural change in (a subset of) the parameters. Generalizations to multiple breaks are possible, but typically computer intensive unless the structure of the model is sufficiently simple; see, e. g. Bai and Perron (2003). The tests against a structural break alternative are based on partial-sample GMM (PS-GMM) estimators and can be of the supremum Wald, Lagrange multiplier (LM), and likelihood ratio (LR) type. Modifications of these tests that use weighted averages rather than the supremum of the tests over all possible break points are proposed by Ploberger et al. (1989) and Andrews and Ploberger (1994). To be consistent in our exposition, we focus on the LM based version of the test. This precludes the need to estimate a possibly highly non-linear model over many different subsamples, which can be costly in terms of computational time, particularly during the exploratory modeling phase.

Let $\pi \subset (0,1)$ and let $\lfloor \pi T \rfloor + 1$ denote the breakpoint of the parameter f_t, where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}$. The null and alternative hypothesis for the Andrews test are given by

$$\mathrm{H}_0 \;:\; f_t = \bar{f}_0 \;\; \forall t \geq 1 \text{ and some } \bar{f}_0 \in F \subset \mathbb{R}^k, \tag{8}$$

$$\mathrm{H}_1 \;:\; \bigcup_{\pi \in \Pi} \mathrm{H}_{1,T}(\pi) \text{ for some } \Pi \subset (0,1), \tag{9}$$

$$\mathrm{H}_{1,T}(\pi) \;:\; f_t = \begin{cases} \bar{f}_1(\pi) & \text{for } t = 1, \ldots, \lfloor \pi T \rfloor \\ \bar{f}_2(\pi) & \text{for } t = \lfloor \pi T \rfloor + 1, \ldots, T, \end{cases} \tag{10}$$

for constants $\bar{f}_1(\pi), \bar{f}_2(\pi) \in F$. Though the test is designed for a single structural break at unknown date, it is by now well-known that the test also has good power properties against a range of other, more general alternatives; see for example the survey of Hansen (2001).

The distribution of the Andrews' LM test is that of the supremum of the square of a tied down Bessel process as derived in Theorem 3 of Andrews (1993), where one can also find the

critical values of the test.

In contrast to the tests described in Sections 2.1 and 2.2, the Andrews' LM test does not build on the autocorrelation of the score of the likelihood, but rather on the average level of the score before and after the break. In particular, the crucial ingredients of the test are scaled versions of $\sum_{t=1}^{\lfloor \pi T \rfloor} \nabla_{\omega,t}$ and $\sum_{t=\lfloor \pi T \rfloor+1}^{T} \nabla_{\omega,t}$. This clearly follows from the framework for the alternative, which in this case is that of a structural break (at unknown point). It also follows directly that if there are regular switches between alternative values of the parameter, the Andrews' test may have difficulty in picking this up: the means of the moment conditions before and after any particular tentative breakpoint may fail to be sufficiently different in small samples. Such cases might arise if the time-varying parameter follows a data generating process with regular regime switches or with strongly mean reverting dynamics.

2.4 Martingale type time–variation

Our final benchmark is the all-purpose test for parameter of Nyblom (1989). This test is based on the assumption that under the alternative the time varying parameter follows a martingale process. As argued by Nyblom (1989), this encompasses the case of one or more structural breaks. The test thus has links to both tests in Sections 2.2 and 2.3.

The key element in the Nyblom test is the partial sum of the likelihood scores. This brings the test close to the partial sums in the Andrews' test. However, the Nyblom test does not take a supremum, but rather considers the average of the squares of partial sums. As we see in our simulation experiment in the next section, this causes the Nyblom test to have an inferior performance compared to the other three tests in most settings.

3 Monte Carlo study

3.1 Set-up

We consider a range of different data generating processes (DGPs). For each DGP, we generate a time series of length $T = 2,000$ observations and compute the GAS-$LM(1)$, the GAS-$LM(5)$, the sup $-LM$ test of Andrews (1993), the test of Müller and Petalas (2010) (MP10), and the test of Nyblom (1989). All test results are stored, and the process is repeated $N = 10,000$ times to compute the size and power properties of the tests. Together, these tests cover parameter stability tests against a wide range of alternative modeling frameworks.

We differentiate the DGPs considered for this study along two dimensions. First, we consider DGPs with different types of dynamics for the time-varying parameter. In particular, we

consider regime switching type models, models with random structural breaks, and state-space models. Second, we consider different degrees of non-linearities in the parameters of the DGPs, such as time variation in the mean, in the variance, in the dependence structure, and in higher order moments.

3.1.1 Types of time variation

We consider the following different types of DGPs to generate time-varying parameters.

Regime switches: Let $n_b \in \mathbb{N}$ denote the fixed number of switches, then the evolution of f_t is given by

$$f_t = \begin{cases} \Delta & \text{for } \lfloor \frac{j \cdot T}{n_b+1} \rfloor + 1 \leq t \leq \lfloor \frac{(j+1) \cdot T}{n_b+1} \rfloor \text{ for every } j = 1, 3, \ldots, (2 \cdot \lfloor 0.5 n_b - 0.5 \rfloor + 1), \\ 0 & \text{otherwise,} \end{cases} \tag{11}$$

with Δ denoting the difference (in absolute value) between the two regimes. For example, for $n_b = 4$, we have $f_t = \Delta$ for $\lfloor 0.2T \rfloor + 1, \ldots, \lfloor 0.4T \rfloor$, and for $\lfloor 0.6T \rfloor + 1, \ldots, \lfloor 0.8T \rfloor$, and zero elsewhere. This creates regular and equally sized patches where the parameter alternately takes the value 0 and Δ. Alternatively, we could make the regime switches stochastic rather than deterministic, but we do not expect major differences with the current deterministic set-up in terms of level and power properties of the different tests.

Random structural breaks: For random structural breaks, we follow the set-up of Elliott and Müller (2006). In particular, we generate n_b uniform random numbers in the interval (0,1), $\pi_1, \ldots, \pi_{n_b}$. The parameter then is a random walk with (infrequent) Gaussian increments at the points $\lfloor \pi_j T \rfloor + 1$ for $j = 1, \ldots, n_b$,

$$f_t = \sum_{j=1}^{n_b} \mathbf{1}_{\{t > \lfloor \pi_j T \rfloor\}} v_j, \tag{12}$$

where $\mathbf{1}_A$ is the indicator function for the event A, and v_j is a Gaussian random variable with zero mean and standard deviation Δ.

State-space: For a DGP with parameter driven dynamics, we assume that f_t follows an autoregressive process of order one, AR(1),

$$f_{t+1} = \phi f_t + \eta_{t+1}, \tag{13}$$

where η_{t+1} is normally distributed with zero mean and standard deviation Δ.

3.1.2 Types of non-linearities in the DGP

For each one of the three different dynamic frameworks for f_t discussed in Section 3.1.1, we consider different models. In particular, we consider models for a time-varying mean, $y_t =$

$f_t + \sigma_\varepsilon \varepsilon_t \in \mathbb{R}$, with $\sigma_\varepsilon \in \mathbb{R}^+$; a time-varying log-variance, $y_t = \exp(f_t/2)\varepsilon_t \in \mathbb{R}$; and time-varying dependence,

$$y_t = \begin{pmatrix} 1 & \tanh(f_t) \\ \tanh(f_t) & 1 \end{pmatrix}^{\frac{1}{2}} \varepsilon_t \in \mathbb{R}^2. \tag{14}$$

We consider two distributions for ε_t, namely the standard normal and the Student's $t(5)$ distribution.

Our fourth model is a time-varying beta distribution $y_t \sim \text{Beta}(\alpha_t, \beta_t)$. We use the model in the empirical application in Section 4 and therefore introduce it more thoroughly. The model has two time-varying parameters. In the DGP, we let both of these depend on a common scalar f_t. In the application later on, we allow both parameters to evolve independently, causing f_t to be two-dimensional.

We consider two settings for the beta DGP. In the first setting, we use

$$\alpha_t = \bar{f} * \frac{\exp(f_t)}{1+\exp(f_t)}, \qquad \beta_t = \frac{\bar{f}}{1+\exp(f_t)}, \qquad \bar{f} > 0, \tag{15}$$

such that both the mean $\mu_t = \exp(f_t)/(1+\exp(f_t))$ and variance $\mu_t(1-\mu_t)/(1+\bar{f})$ of the beta distribution vary over time. The mean lies in the [0,1] range by construction, irrespective of the value of f_t. The variance automatically tends to zero if the mean tends to either 0 or 1, which is natural for the beta distribution. The constant $\bar{f} > 0$ determines the additional extent of concentration of the distribution.

In the second version of the model, we set

$$\alpha_t = \bar{f} * \exp(f_t), \qquad \beta_t = (1-\bar{f}) * \exp(f_t), \qquad 0 < \bar{f} < 1. \tag{16}$$

This implies that the mean $\mu_t = \bar{f}$ is constant, while the variance $\bar{f}(1-\bar{f})/(1+\exp(f_t))$ varies over time.

In total we have models for means, variances, and correlations for the normal and Student's t distribution, plus two versions of the time varying beta distribution. Combining these 8 settings in total with the 3 different forms of time-variation from Section 3.1.1, we have 24 simulation experiments in total. For each of these, we implement all tests at the 5% significance level.

Before we present the results, we note that the Andrews test is implemented over a grid of breakpoints Π in (9). We use the test based on the boundary break point values of 15% and 85% of the sample size. Ideally, we would compute the test using all possible breakpoints between these boundary values. This, however, is too computer intensive, certainly for the non-linear data generating processes considered later on. The supremum over a grid of values is always less than the supremum over the entire set of possible breakpoints. This entails the danger of the test in the simulations are undersized and their power is too low. To avoid this

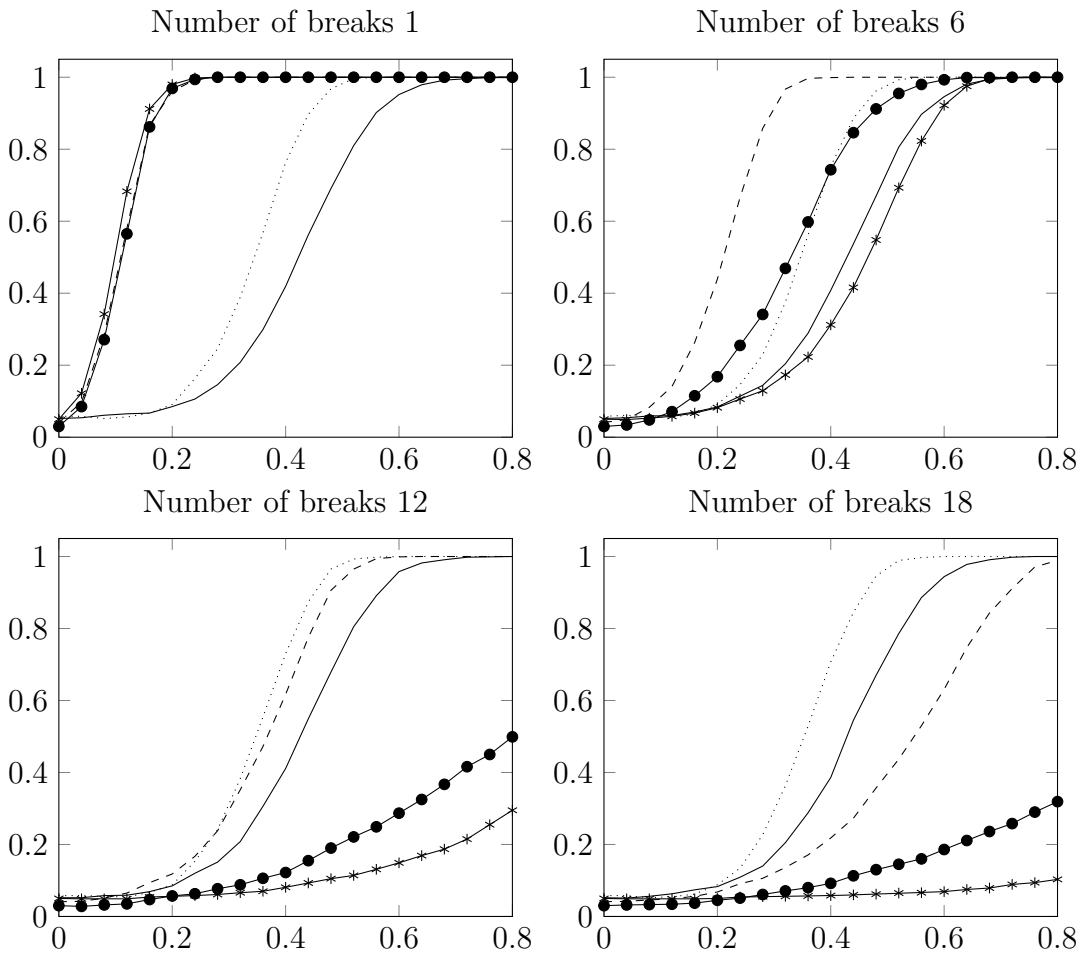

Figure 1: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for the regime–switching Gaussian time-varying mean model.

problem we calibrated the grid such that the simulated size of the test is roughly equal to the nominal size given the critical value in Andrews (1993). The final grid of breakpoint values is $\lfloor jT \rfloor$ for $j = 0.15, \ldots, 0.85$ in steps of 0.0005.

3.2 Results

The results for the time-varying mean, Gaussian distribution, and deterministic regime switches are presented in Figure 1. For the case of a single regime switch, the top left panel shows that the power of the Nyblom, Andrews, and Müller-Petalas is best. This is to be expected, as the Andrews test is the optimal test in this case. The power behavior of all three tests roughly coincides. The GAS tests are less powerful and need roughly a 2.5 to 3 times more distant alternative than the Andrews test to obtain maximum power.

If the number of regime switches increases, the Nyblom test quickly looses power, followed by the Andrews sup-LM test. Already for 6 regime switches over 2,000 observations, the Nyblom test has similar power to the GAS(0,1) test, whereas the Andrews test has similar

power to the GAS(0,5) test. The MP10 test's power also decreases, but the MP10 test is still the best for 6 regime switches. Further increases in the number of switches, however, also breaks down the performance of the MP10 test. The power performance of the GAS tests, on the other hand, remains remarkably robust across the number of regime switches. Although this may seem suspicious at first sight, the result is entirely intuitive. The GAS test is based on the autocorrelation of the likelihood score with respect to the dynamic parameter, where the scores are computed under the null. Looking at Figure 2, the parameter estimate under the null is some type of average level of the true parameter path, which is indicated by the pulse function in each graph. As a result, the scores under the null roughly follow the pattern of the (demeaned) pulse function. This exhibits strong autocorrelation. If the number of regime switches increases, the autocorrelation remains strong. The number of points where the correlation pattern is broken, is equal to the number of switches. As the latter is typically small compared to the sample size, the power performance of the GAS test remains highly stable if we increase the number of switches.

The behavior is very different for the MP10 test. For this test, the estimated difference $f_t - \bar{f}$ under the alternative $c = 10$ plays a key role, as explained in Section 2.2. Figure 2 also holds the MP10 estimate of f_t. It is clear that the increase in the number of regime switches makes it harder for the smoothed estimate $\hat{f}_t$ to capture the true dynamics of the simulated parameter. As the number of regime switches increases, the estimated difference $f_t - \bar{f}$ becomes negligible. As a result, the power of the MP10 tests starts to decrease.

The results for the case with random breaks are given in Figure 11. For one break, the tests do not appear to reach maximum power of one. This is due to the fact that the generated break dates are sometimes very close to the starting or end point of the sample. The tests have little power against these alternatives. If the number of regime switches increases, maximum power is reached quite fast. This is due to the substantial probability that two consecutive breaks will be in the same direction, thus increasing the overall signal that the parameters are not constant over time. In all cases, the Andrews, MP10, and Nyblom tests appear to have superior power to the GAS based tests.

Finally, Figure 12 contains the results for the parameter driven time variation. The time-varying parameter follows the model $f_{t+1} = \phi f_t + \sigma_\eta \eta_t$, where $\eta_t \sim \mathrm{N}(0, 1)$. The left hand panel contains the results for varying σ_η^2 on the horizontal axis, and fixed $\phi = 0.9$. The right hand panel contains the results for varying ϕ and fixed $\sigma_\eta^2 = 0.15$.

In both settings, the GAS tests display the best overall power performance. For the case of fixed $\phi = 0.9$, the true simulated parameter path exhibits strong mean reversion. In a sense, this is similar to the regular regime switches in Figure 10. We already explained why the

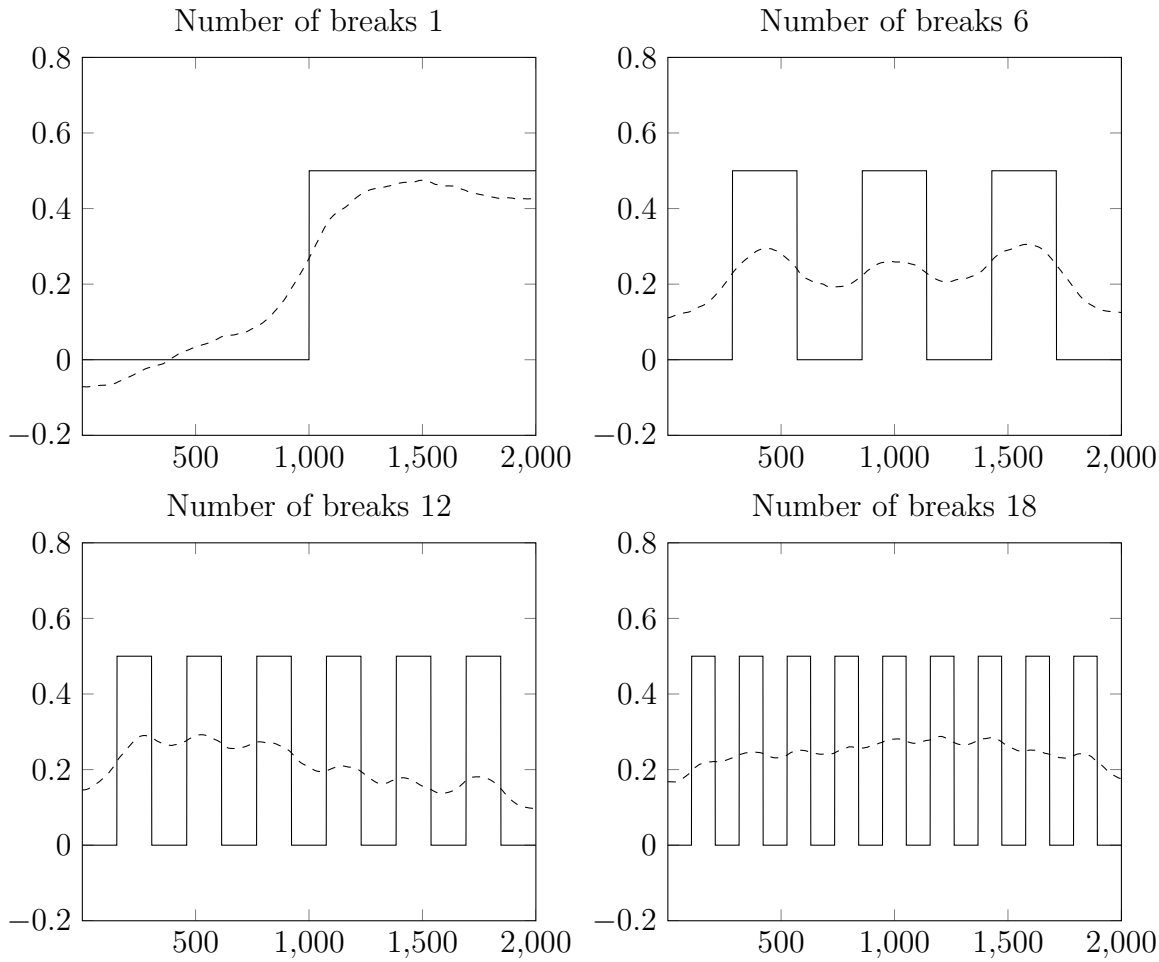

Figure 2: Evolution of the parameter path estimated using the Müller-Petalas procedure with $c = 10$ (dashed line) and the true (simulated) parameter path (solid line). The different panels contain the results for an increasing number of regime switches.

MP10 test in this case has worse power performance compared to the GAS tests. The same phenomenon is at work in the left panel in Figure 12. What is difficult to see in the current figure, but will be made clear later on is that for local alternatives, i.e., values of σ_η^2 very close to zero, the MP10 test actually has a better power performance than the GAS test. This is in line with the analytical results of Müller and Petalas (2010), who prove that their test is optimal against local alternatives.

In the right panel of Figure 12, we again see that the GAS tests have the best performance. For large values of ϕ (on the horizontal axis), the time variation becomes a martingale, and the Nyblom, Andrews, and MP10 test display adequate power behavior. For lower values of ϕ, however, the parameter path is strongly mean reverting and only the GAS test has power. The power comes from the fact that even for strong mean reversion, the score under the null still displays significant autocorrelation. There is, however, no substantial change in the level of the score under the null (Andrews and Nyblom) for small values of ϕ, nor is there a strong

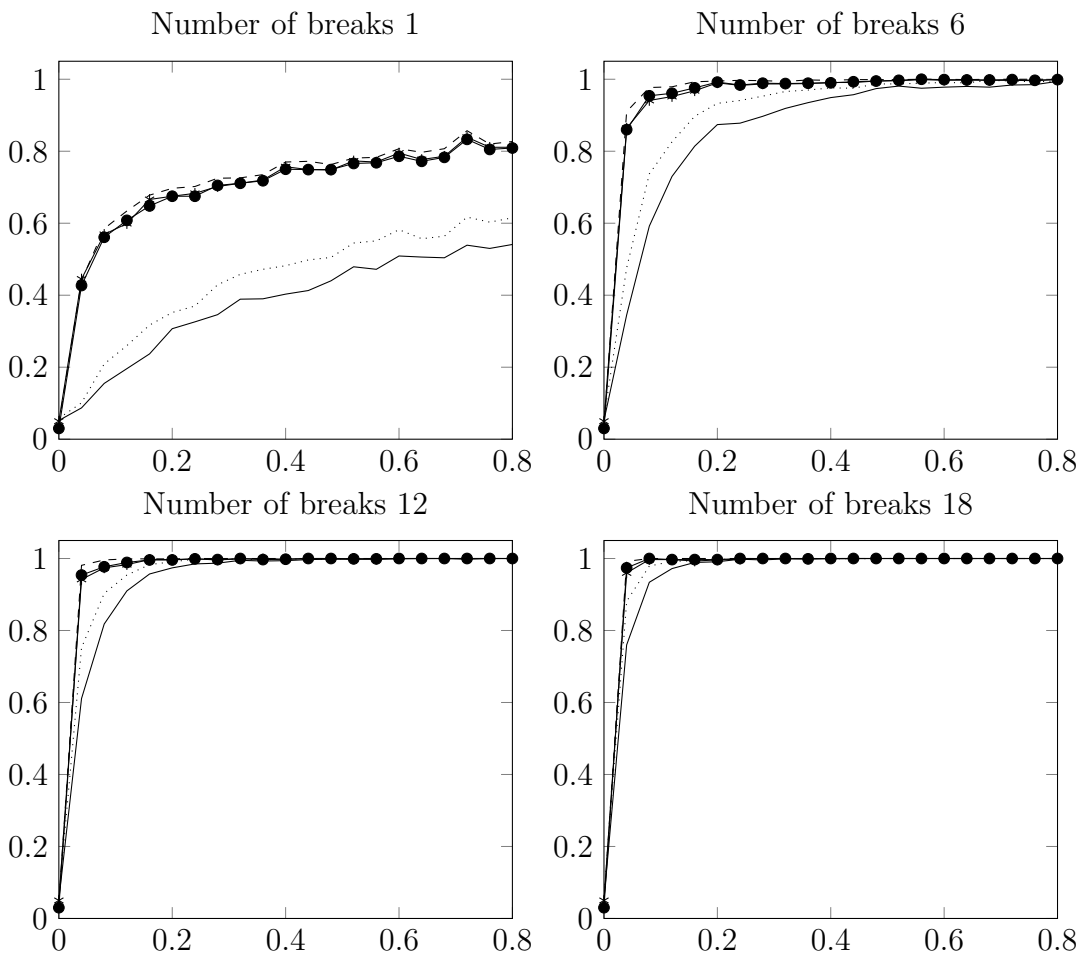

Figure 3: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for the Gaussian time-varying mean model with random breaks.

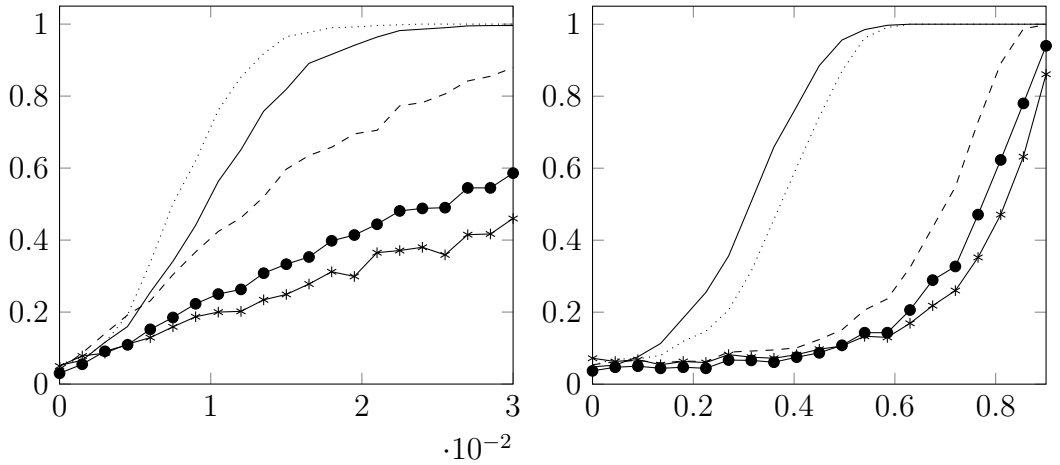

Figure 4: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for the Gaussian time-varying mean model with parameter driven time variation.

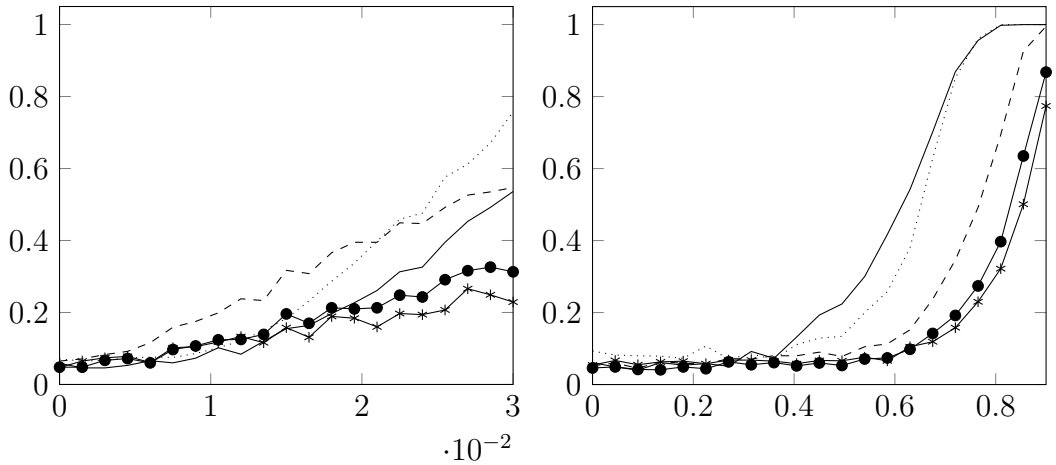

Figure 5: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for the time-varying beta model with parameter driven time variation.

time variation in $\hat{f}_t - \hat{\bar{f}}$ in this case, compare Figure 2.

Rather than presenting the full simulation results in the main text, we refer the interested reader to the online Appendix containing all additional results. The results are highly robust compared to the results presented so far. This holds if we replace the Gaussian distribution by a Student's t distribution, or if we consider time-varying variances and correlations rather than time-varying means.

Also the results for the time-varying beta distribution confirm our earlier findings. We highlight, however, one main finding. In Figure 5, we present the results for the case of a time-varying mean and variance, i.e., the first simulation setting for the time-varying beta distribution. The time-varying parameter f_t follows an autoregressive process of order 1. We particularly highlight the left hand panel in Figure 5. The variance σ_η^2 of the error term is in this case such that we remain close to the null hypothesis of no time variation. As a result, the maximum power over the range of alternatives considered remains quite low and we effectively zoom in on the local power behavior of all tests. The left panel clearly shows the MP10 test has the best power very close to the null. As mentioned earlier, this is directly in line with the analytical results of Müller and Petalas (2010). As we get more distant from the null, however, the GAS tests have a better performance than the MP10 test.

Also the right hand panel in the figure displays a similar pattern as in Figure 12. For medium persistence levels, the GAS tests clearly outperform the other tests. For low persistence levels, however, all tests have hardly got any power. If the true persistence level is about 0.7, the GAS tests reach their maximum power level. For the MP10 test this happens for a persistence level of about $\phi = 0.9$. In the next section, we implement all tests for the time-varying beta model

to an empirical example and provide some further insight.

4 Empirical application

We consider quarterly observations of losses given default (LGDs) on corporate bonds. The data are obtained from Moody's and cover the first quarter of 1982 to the first quarter of 2010. The fraction of loss is measured as the percentage price drop in the value of the corporate bond from the day before to 20 days after the announcement of default.[1] We want to test whether there is significant time variation in the distributional characteristics of LGDs. Time varying LGDs are important for credit risk modeling and financial risk management, as credit portfolio losses could be severely underestimated if default risk and LGD risk exacerbate one another; see for example Creal et al. (2014).

The data display several non-standard features. First, because LGDs are measured as percentage losses, they are bounded to the interval $[0, 1]$. To accommodate this, we assume that the LGDs are drawn from a beta distribution. Second, the number of observed LGDs varies per quarter, such that the dimension of the observation vector varies over time. These features have to be accounted for in the testing methodology. Combining the observation period with the varying number of LGDs per quarter, we have 1125 LGD-quarter observations, with the number of LGDs per quarter varying from 1 in 1982 to a max of 58 in 2009.

Let $y_{i,t}$ denote the ith observation at time t with $i = 1, \ldots, K_t$, where K_t represents the number of LGD observations at time t. We take K_t as given and model $y_{i,t}$ at time t as independent draws from a beta distribution with parameters $\alpha_t = \exp(f_{1,t})$ and $\beta_t = \exp(f_{2,t})$, where $f_t = (f_{1,t}, f_{2,t})'$. Define $y_t = (y_{1,t}, \ldots, y_{K_t,t})'$. Then the log conditional observation density of y_t is given by

$$\ln p(y_t|f_t) = \sum_{i=1}^{K_t} \ln\Gamma(\alpha_t + \beta_t) - \ln\Gamma(\alpha_t) - \ln\Gamma(\beta_t) + (\alpha_t - 1)\ln y_{i,t} + (\beta_t - 1)\ln(1 - y_{i,t}), \quad (17)$$

where Γ denotes the gamma function. The conditional score and information matrix for (17) are given by

$$\nabla_{f,t} = \sum_{i=1}^{K_t} \begin{pmatrix} \big(\Psi(\alpha_t + \beta_t) - \Psi(\alpha_t) + \ln y_{i,t}\big) \cdot \alpha_t \\ \big(\Psi(\alpha_t + \beta_t) - \Psi(\beta_t) + \ln(1 - y_{i,t})\big) \cdot \beta_t \end{pmatrix}, \quad (18)$$

[1] This is also known as the market implied LGD. As the value may become negative, e.g., after a timely restructuring or merger announcement, we censor negative observed LGDs to 1 basis point, i.e., 0.01%. This concerns only 14 out of 1125 observations, or 1.25% of the observations.

Table 1: Test statistics and critical values for the corporate LGD data, 1982Q1–2010Q1

	Stat	10%	5%	1%
LM GAS(0,1)	16.20	4.60	5.99	9.21
LM GAS(0,5)	22.84	15.99	18.31	23.21
MP10	-27.34	-12.80	-14.32	-17.57
MP10*	-40.78	-12.80	-14.32	-17.57
Andrews	18.56	10.01	11.79	15.51
Nyblom	1.63	0.61	0.75	1.07

and

$$\mathcal{I}_t = K_t \cdot \begin{pmatrix} \alpha_t^2(\Psi'(\alpha_t) + \Psi'(\alpha_t + \beta_t)) & -\alpha_t\beta_t\Psi'(\alpha_t + \beta_t) \\ -\alpha_t\beta_t\Psi'(\alpha_t + \beta_t) & \beta_t^2(\Psi'(\beta_t) + \Psi'(\alpha_t + \beta_t)) \end{pmatrix}, \tag{19}$$

where Ψ denotes the digamma function, i.e., $\Psi(x) = \mathrm{d}\ \ln\Gamma(x)/\mathrm{d}\,x$. We set the GAS scaling matrix to the inverse information matrix, i.e., $S_t = \mathcal{I}_t^{-1}$, to account for the curvature of the score. Using these definitions, we can now compute the different test statistics. The results are presented in Table 1.

All test statistics clearly reject the null hypothesis of constant parameters. We have slightly modified the Muller-Petalas test (denoted as MP10*) to account for the fact that the number of observations K_t varies over time. In the original MP10 paper, the Hessian is estimated unconditionally over the entire sample. That makes sense if the number of observations per period is constant. Here, however, we treat K_t as given and multiply the Hessian in the algorithm of MP10 at time t by $K_t/\bar{K}$, with K the time series average of K_t. The modification result follows directly from a similar derivation as used for the information matrix in (19), and corrects the steps in the MP10 algorithm for periods where there are many, respectively few LGD observations in the cross section.

We confirm the test results from Table 1 by estimating the path of the time varying parameter f_t in two alternative ways. First, we estimate f_t based on a GAS(1,1) model,

$$f_{t+1} = \omega + As_t + Bf_t, \quad s_t = S_t\nabla_t, \quad A = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}. \tag{20}$$

Estimating the model with full rather than diagonal matrices A and B gives very similar results. The parameter estimates are presented in Table 2. We find that there is strong persistence in both α_t and β_t, as both b_{11} and b_{22} are large. All parameter estimates are strongly significant, except ω_2. Interestingly, the persistence (b_{11}) in α_t is not as strong as that in β_t. As α_t and β_t characterize the mean of the beta distribution when it is close to 0 and 1, respectively,

Table 2: GAS(1, 1) coefficients estimation results

Coeff.	Estimate	Std Err	t–stat
ω_1	0.0559	0.0088	6.3449
ω_2	-0.0023	0.0073	-0.3075
a_{11}	0.1943	0.0220	8.8156
a_{22}	0.1836	0.0361	5.0770
b_{11}	0.8571	0.0233	36.8217
b_{22}	0.9235	0.0355	26.0252

the higher persistence (b_{22}) of β_t indicates that the higher LGDs are more persistent than low LGDs. There appears to be no such difference between a_{11} and a_{22}.

Our second estimate of f_t is obtained as a byproduct of the MP10 algorithm. It is based on a Weighted Average Risk estimate of the path f_t for several local alternatives. We use the same method as for MP10* to correct for the time-varying number of observations K_t when estimating the path. The results are shown in Figure 6.

The LGD data range from close to zero to almost one for given cross sections. Rather than providing the plots of α_t and β_t directly, we present a plot of the mean of the beta distribution $\alpha_t/(\alpha_t + \beta_t)$. Both the MP10* and GAS estimates of the mean capture the salient features of the data. There are clear peaks in average credit losses around the 1991 recession, the 2000-2001 burst of the dotcom bubble, and the most recent financial crisis. The peaks clearly defy the assumption of constant parameters. The MP estimate appears to lead the GAS estimate. This is due to the fact that the GAS model is a filter (one-sided estimator), whereas the MP estimate is a smoother (two-sided estimator) and thus takes future observations into account. There is one episode where the two estimates differ substantially, namely the period leading up to the 2008 financial crisis. The GAS estimate reveals a more moderate trough than the MP estimate.

It is interesting to see that the MP estimate works well even though it is designed for *local* time variation only. To understand this, note that the smoothed path of f_t in Figure 6 for MP10* is based on a weighted average of 10 different paths, corresponding to autoregressive coefficients $b_{11} = b_{22} = 1 - c/T$ for $c = 0, 5, 10, \ldots, 50$. If we look closer to the weights, we find that all the weight is assigned to paths corresponding to $c = 30, \ldots, 50$, with the mode weight at $c = 40$. As $T = 113$, this corresponds to an autoregressive decay of $1 - 40/113 \approx 0.65$. This is much lower than the persistence estimates of the GAS model in Table 2. Moreover, the autoregressive coefficient in the MP10 method is the same for α_t and β_t, in contrast to the GAS model.

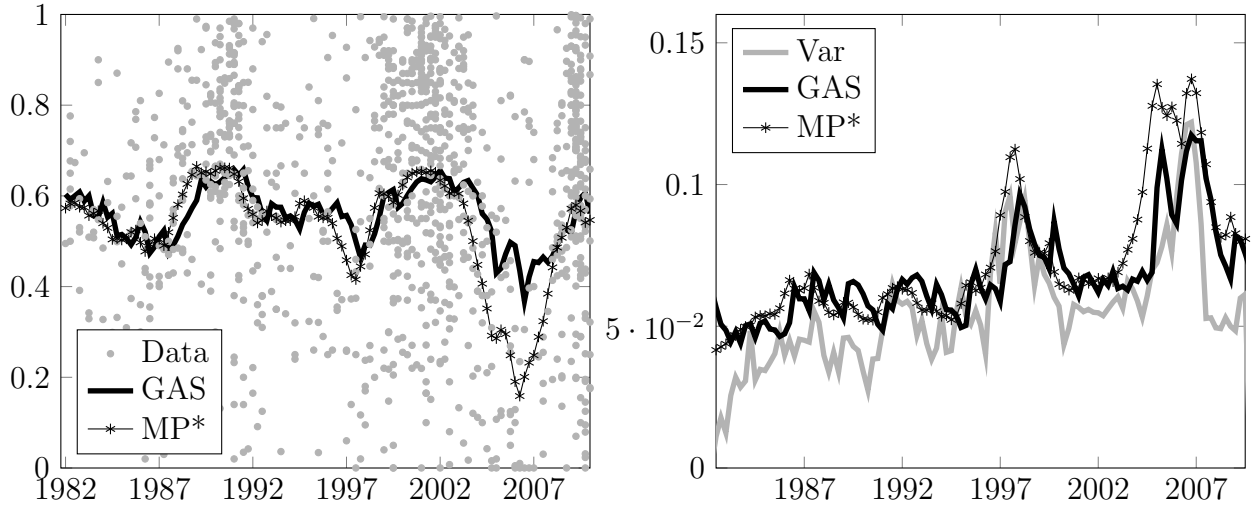

Figure 6: Corporate Loss Given Default (LGD) data, 1982–2010, and fitted means (left panel) and variances (right panel)

Note: This figure contains the market implied LGDs of corporate bonds over the period 1982Q1–2010Q1 as observed by Moody's, left panel. The left panel also contains the mean of the fitted beta distribution, $\alpha_t/(\alpha_t+\beta_t)$, for the GAS model from Table 2 and the MP10* smoothed parameter path of Müller and Petalas (2010). The right hand curve provides the estimates of the variance, $\alpha_t/((\alpha_t+\beta_t)^2(1+\alpha_t+\beta_t))$, for both methods, as well as a 1 year rolling window estimate of the variance (Var).

To fit the unconditional variation in the data, the smaller persistence parameter in MP10* is counterbalanced by a higher error variance. Combined, these two effects make the MP10* path more sensitive to some of the lower values of the LGDs in the period leading up to the credit crisis. Also note that the MP10* test is not influenced by these less persistent paths, as it is entirely based on the local alternative $c = 10$ or a persistence parameter $1 - 10/113 \approx 0.91$. The latter is more in line with the estimated persistence parameters b_{11} and b_{22} of the GAS model.

We present estimates of the variance of the beta distribution in the right-hand panel in Figure 6. The variance equals $\alpha_t/((\alpha_t+\beta_t)^2(1+\alpha_t+\beta_t))$ and is slightly trending upwards, with the more recent variation in LGD percentages being somewhat larger than that in the early 1980s. There are two peaks in the variance. These are linked to the periods when there are few LGD observations and the relative dispersion of the few observed LGDs is large. Again, the variance estimates using either the MP10* or the GAS framework are roughly similar. The main differences arise around 1997 and around 2004–2006. The lower mean over the latter period for MP10* compared to GAS (see left panel) is thus partly off-set by a higher variance. Again, the test is not affected by this, because it uses an autoregressive coefficient of around 0.91 for MP10* ($c = 10$), rather than around 0.65 ($c = 40$) as used to obtain the smoothed estimate.

5 Conclusions

We have introduced a new mis-specification test for parameter stability in general non-linear, possibly non-Gaussian time-series models. Building on the Generalized Autoregressive Score (GAS) dynamics as developed by Creal et al. (2013), we proposed a Lagrange Multiplier test for the null of constant parameters against the alternative of GAS effects. We have carried out an extensive Monte Carlo study to investigate the finite sample properties of the new test compared to a number of competing general purpose tests for parameter instability. Each of these tests against a very different time-varying parameter framework: either structural breaks, Andrews (1993); local parameter driven variation, Müller and Petalas (2010); or martingale processes for the time-varying parameters, Nyblom (1989).

We find that the new GAS test has a robust power performance. For different types of DGPs, the power behavior of the GAS test remains relatively constant, whereas that of competing tests varies considerably. None of the tests is uniformly superior in all situations considered. The GAS test performs well if parameters vary considerably over time, particularly when this variation is strongly mean reverting and frequent. For incidental changes or a small magnitude of the time variation, the test of Müller and Petalas (2010) typically performs best, which is in line with what theory predicts.

We applied our tests to an empirical panel data set consisting of loss given default percentages of corporate bonds. We showed how all tests could be used for this data set and could be adapted for a setting with a time-varying number of observations. Interestingly, we found that the smoothing approach of Müller and Petalas (2010) can also be useful in cases of non-local time variation in the parameters. The estimated path using their algorithm gives comparable results to estimating the path using a dynamic GAS model for the data set considered. This again illustrates that the two testing paradigms can provide complementary signals in empirical work.

References

Andrews, D. W. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. Econometrica: Journal of the Econometric Society, 817–858.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. Econometrica 61(4), 821–856.

Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under the alternative. Econometrica 62(6), 1383–1414.

Bai, J. and P. Perron (2003). Computation and analysis of multiple structural change models. Journal of Applied Econometrics 18(1), 1–22.

Bauwens, L. and D. Veredas (2004). The stochastic conditional duration model: A latent factor model for the analysis of financial durations. Journal of Econometrics 119(2), 381–412.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31(3), 307–327.

Chow, G. C. (1960). Tests of equality between sets of coefficients in two linear regressions. Econometrica 28(3), 591–605.

Cipollini, F., R. F. Engle, and G. M. Gallo (2012). Semiparametric vector MEM. Journal of Applied Econometrics forthcoming.

Creal, D. D. (2012). A survey of sequential Monte Carlo methods for economics and finance. Econometric Reviews 31(3), 245–296.

Creal, D. D., S. J. Koopman, and A. Lucas (2011). A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. Journal of Business and Economic Statistics 29(4), 552–563.

Creal, D. D., S. J. Koopman, and A. Lucas (2013). Generalized autoregressive score models with applications. Journal of Applied Econometrics 28(5), 777–795.

Creal, D. D., B. Schwaab, S. J. Koopman, and A. Lucas (2014). Observation driven mixed-measurement dynamic factor models with an application to credit risk. The Review of Economics and Statistics.

Davidson, R. and J. G. MacKinnon (1990). Specification tests based on artificial regressions. Journal of the American Statistical Association 85(409), 220–227.

Davis, R. A., W. T. M. Dunsmuir, and S. Streett (2003). Observation driven models for Poisson counts. Biometrika 90(4), 777–790.

De Lira Salvatierra, I. and A. J. Patton (2013). Dynamic copula models and high frequency data. Duke University Working Paper.

Durbin, J. and S. J. Koopman (2000). Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives (with discussion). Journal of the Royal Statistical Society (Series B) 62, 3–56.

Durbin, J. and S. J. Koopman (2012). Time Series Analysis by State Space Methods (2nd ed.). Oxford: Oxford University Press.

Elliott, G. and U. K. Müller (2006). Efficient tests for general persistent time variation in regression coefficients. The Review of Economic Studies 73(4), 907–940.

Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50(4), 987–1007.

Engle, R. F. and G. M. Gallo (2006). A multiple indicators model for volatility using intra-daily data. Journal of Econometrics 131(1-2), 3–27.

Engle, R. F. and J. R. Russell (1998). Autoregressive conditional duration: A new model for volatility using intra-daily data. Econometrica 66(5), 1127–1162.

Franzini, L. and A. C. Harvey (1983). Testing for deterministic trend and seasonal components in time series models. Biometrika 70(3), 673–682.

Hafner, C. and H. Manner (2012). Dynamic stochastic copula models: Estimation, inference and applications. Journal of Applied Econometrics 27(2), 269–295.

Hamilton, J. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57(2), 357–384.

Hansen, B. E. (2001). The new econometrics of structural change: Dating breaks in us labor productivity. Journal of Economic perspectives 15(4), 117–128.

Harvey, A. C. (1989). Forecasting, structural time series models and the Kalman Filter. Cambridge: Cambridge University Press.

Harvey, A. C. (2013). Dynamic Models for Volatility and Heavy Tails: With Applications to Financial and Economic Time Series. Econometric Series Monographs. Cambridge University Press.

King, M. L. and G. H. Hillier (1985). Locally best invariant tests of the error covariance matrix of the linear regression model. Journal of the Royal Statistical Society. Series B (Methodological) 47(1), 98–102.

Koopman, S. J., A. Lucas, and M. Scharth (2012). Predicting time-varying parameters with parameter-driven and observation-driven models. Tinbergen Institute Discussion Paper 12-020/4.

Lee, J. (1991). A Lagrange multiplier test for GARCH models. Economics Letters 37(3), 265–271.

Müller, U. K. and P. E. Petalas (2010). Efficient estimation of the parameter path in unstable time series models. The Review of Economic Studies 77, 1508–1539.

Nyblom, J. (1989). Testing for the constancy of parameters over time. Journal of the American Statistical Association 84(405), 223–230.

Nyblom, J. and T. Mäkeläinen (1983). Comparisons of tests for the presence of random walk coefficients in a simple linear model. Journal of the American Statistical Association 78(384), 856–864.

Oh, D. H. and A. J. Patton (2013). Time-varying systemic risk: Evidence from a dynamic copula model of cds spreads. Duke University Working Paper.

Patton, A. J. (2006). Modelling asymmetric exchange rate dependence. International Economic Review 47(2), 527–556.

Perron, P. (2006). Dealing with structural breaks. Palgrave Handbook of Econometrics 1, 278–352.

Ploberger, W., W. Krämer, and K. Kontrus (1989). A new test for structural stability in the linear regression model. Journal of Econometrics 40(2), 307–318.

Qu, Z. and P. Perron (2007). Estimating and testing structural changes in multivariate regressions. Econometrica 75(2), 459–502.

Richard, J.-F. and W. Zhang (2007). Efficient high-dimensional importance sampling. Journal of Econometrics 141(2), 1385–1411.

Russell, J. R. (2001). Econometric modeling of multivariate irregularly-spaced high-frequency data. Unpublished manuscript, University of Chicago, Graduate School of Business.

Shephard, N. (2005). Stochastic Volatility: Selected Readings. Oxford: Oxford University Press.

Vogelsang, T. J. and P. Perron (1998). Additional tests for a unit root allowing for a break in the trend function at an unknown time. International Economic Review, 1073–1100.

White, H. (1987). Advances in Econometrics, Fifth World Congress, Volume 1, Chapter Specification testing in dynamic models, pp. 1–58. Cambridge University Press.

Online Appendix: additional simulation results

Gaussian time–varying variance

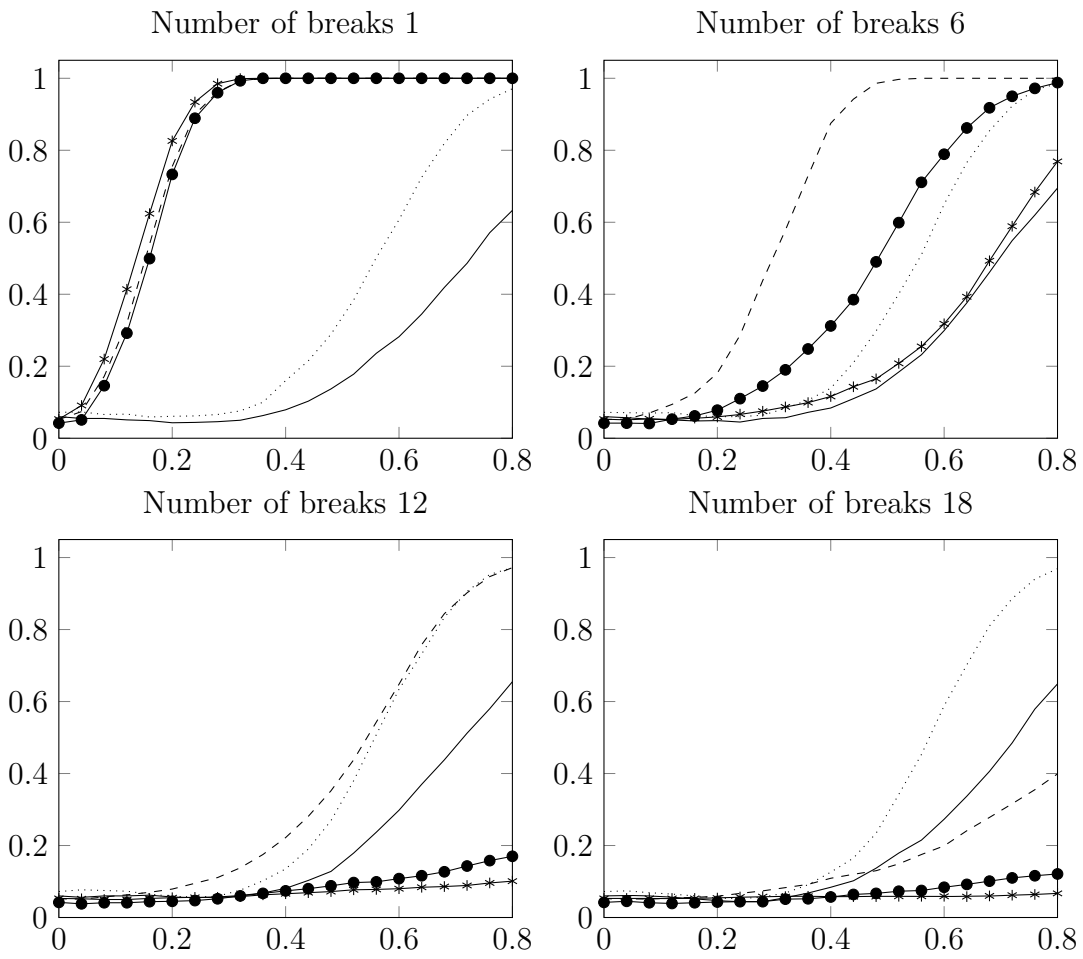

Figure 7: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for regime–switching.

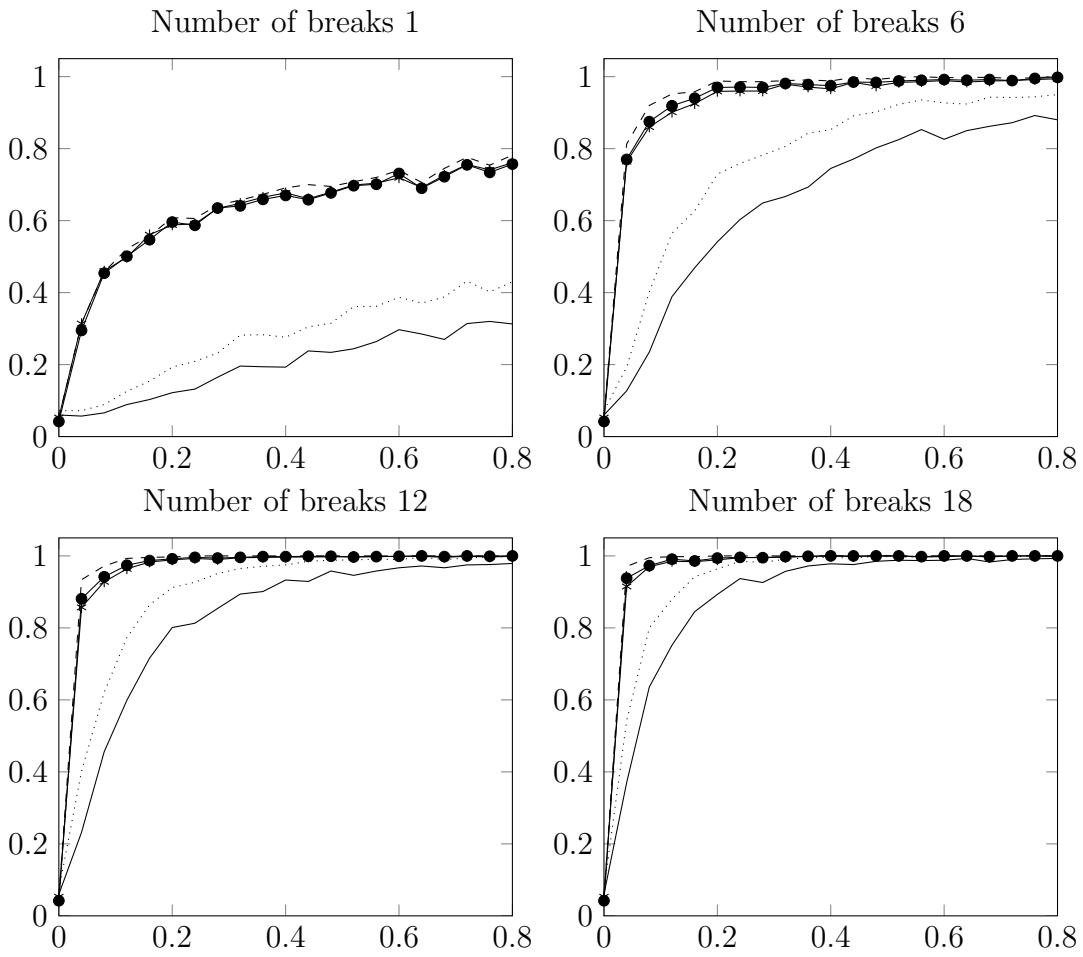

Figure 8: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for random breaks.

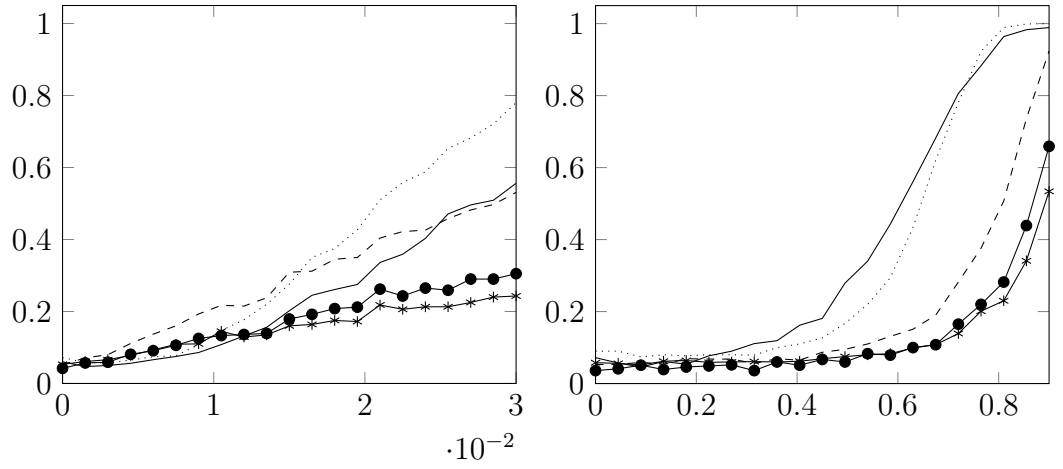

Figure 9: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for state space.

Gaussian time–varying correlation

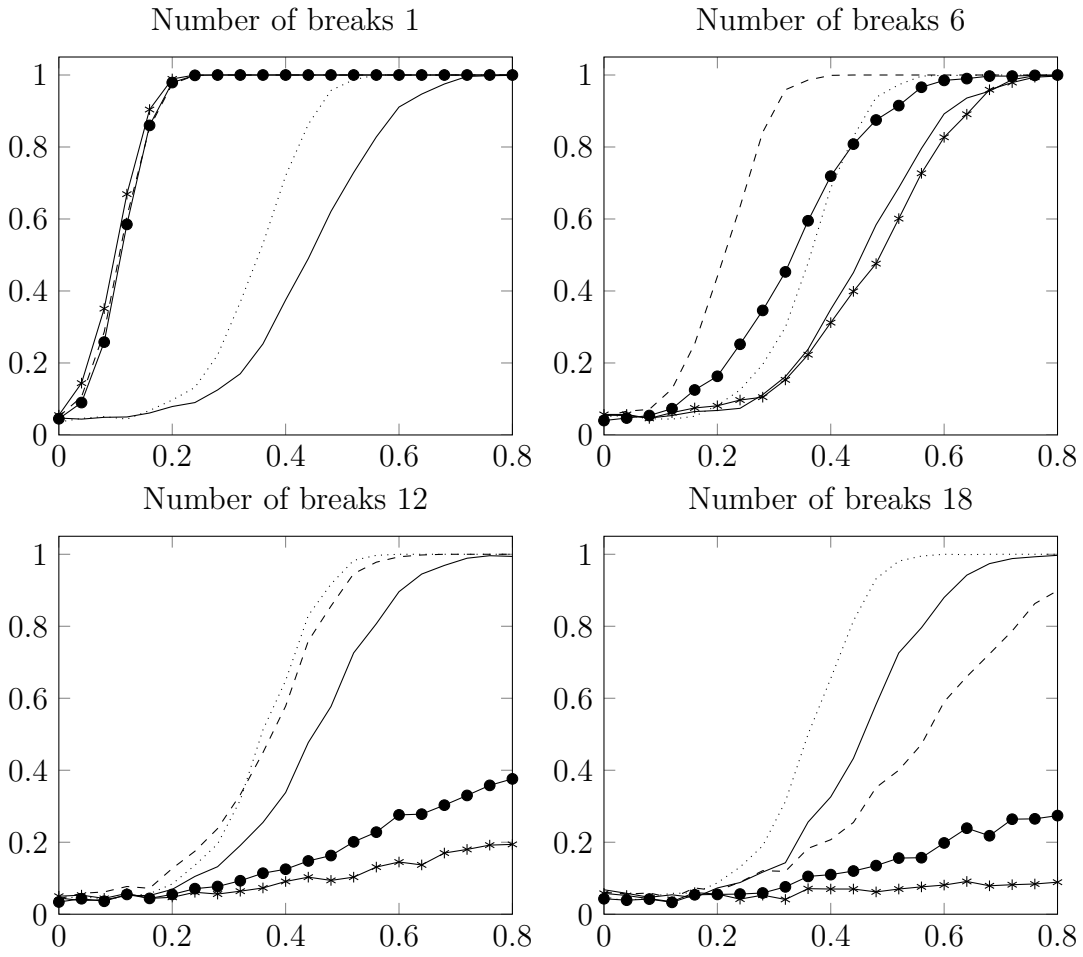

Figure 10: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for regime–switching.

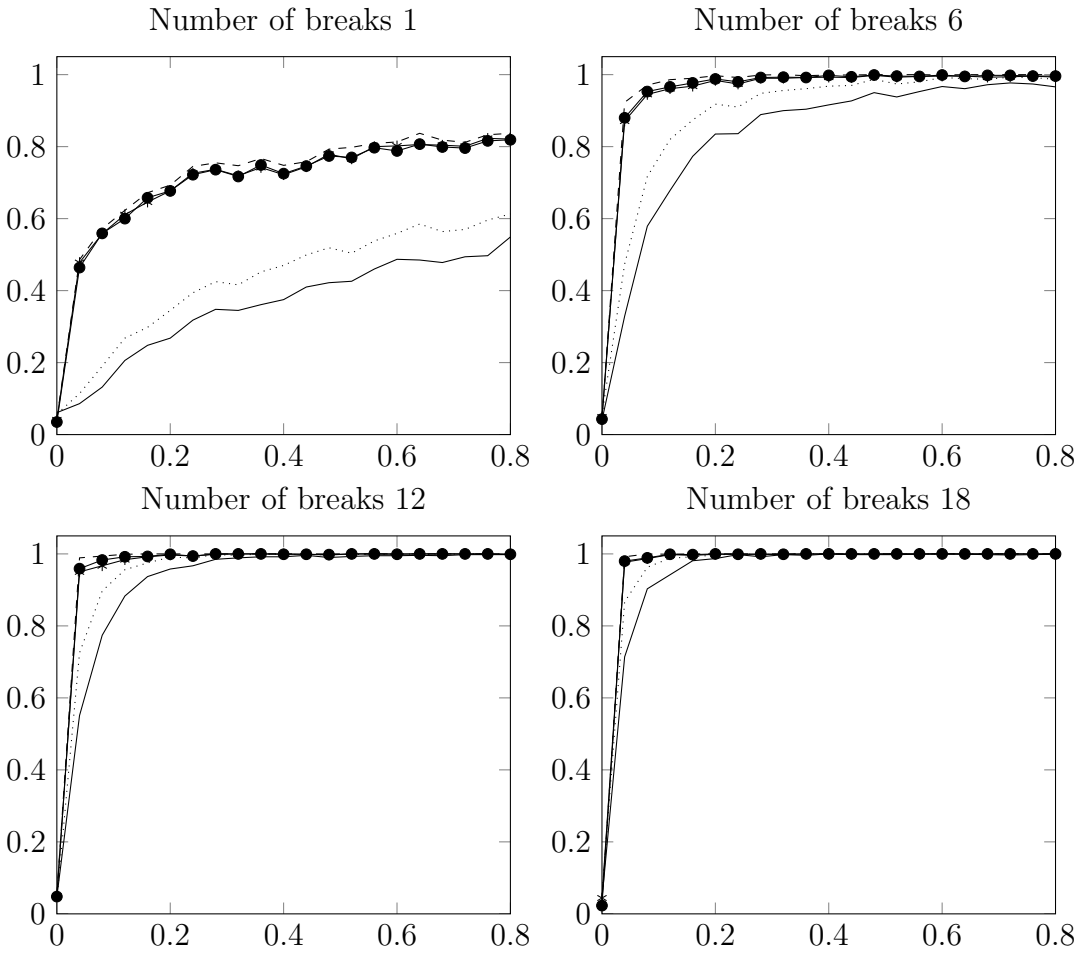

Figure 11: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for random breaks.

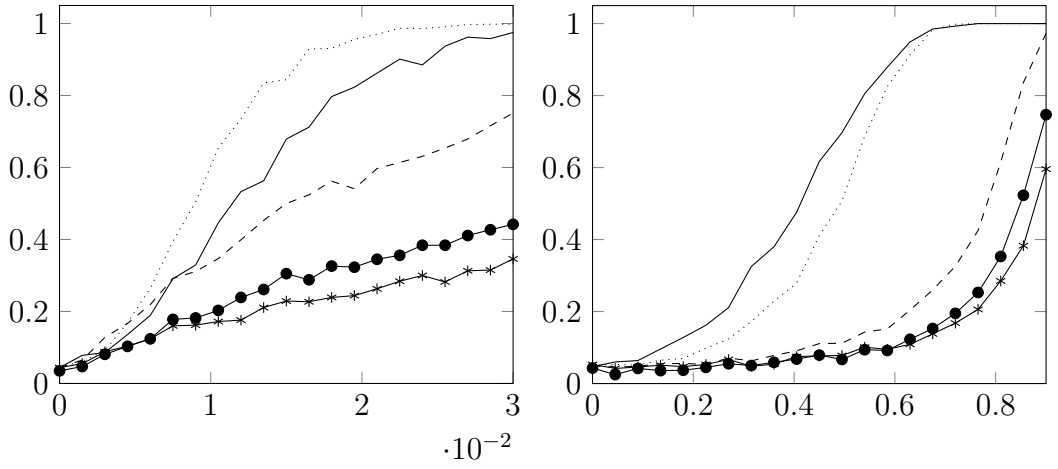

Figure 12: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for state space.

t–distribution time–varying mean

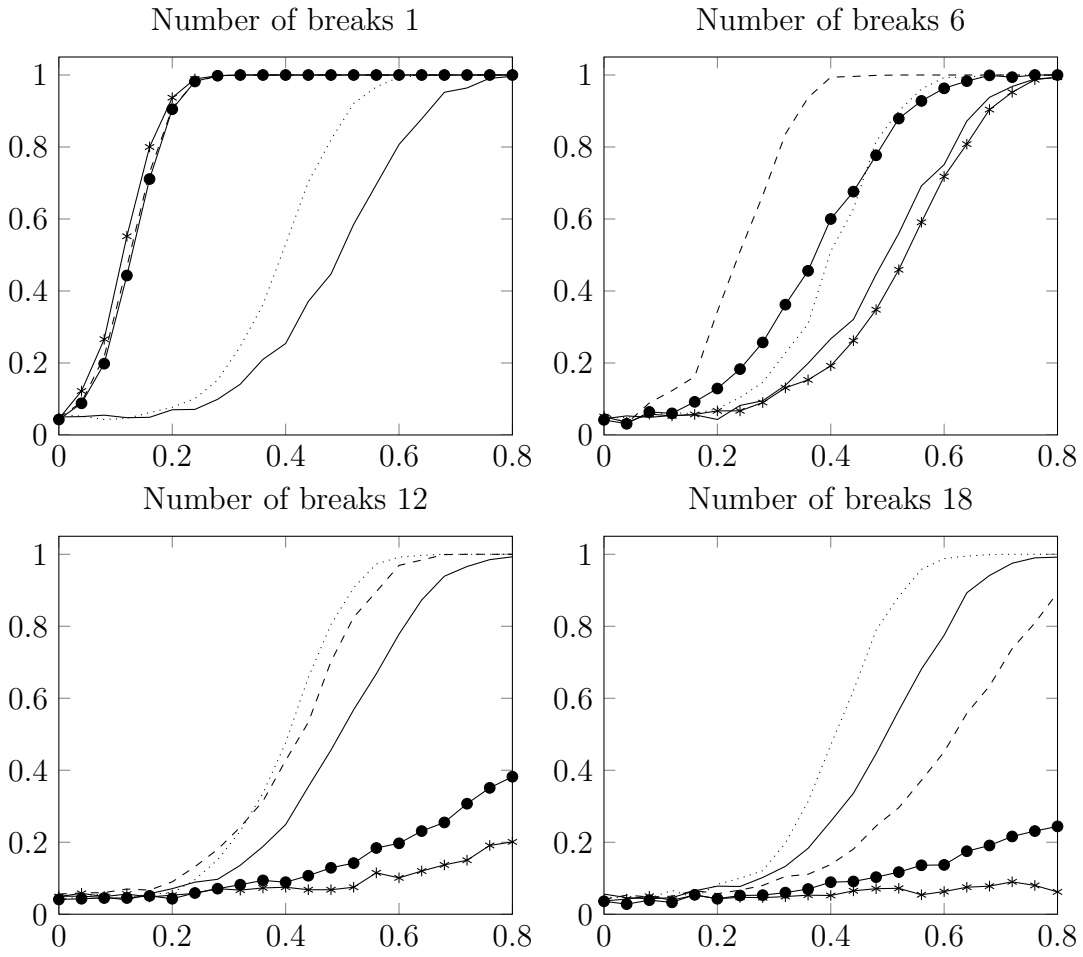

Figure 13: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for regime–switching.

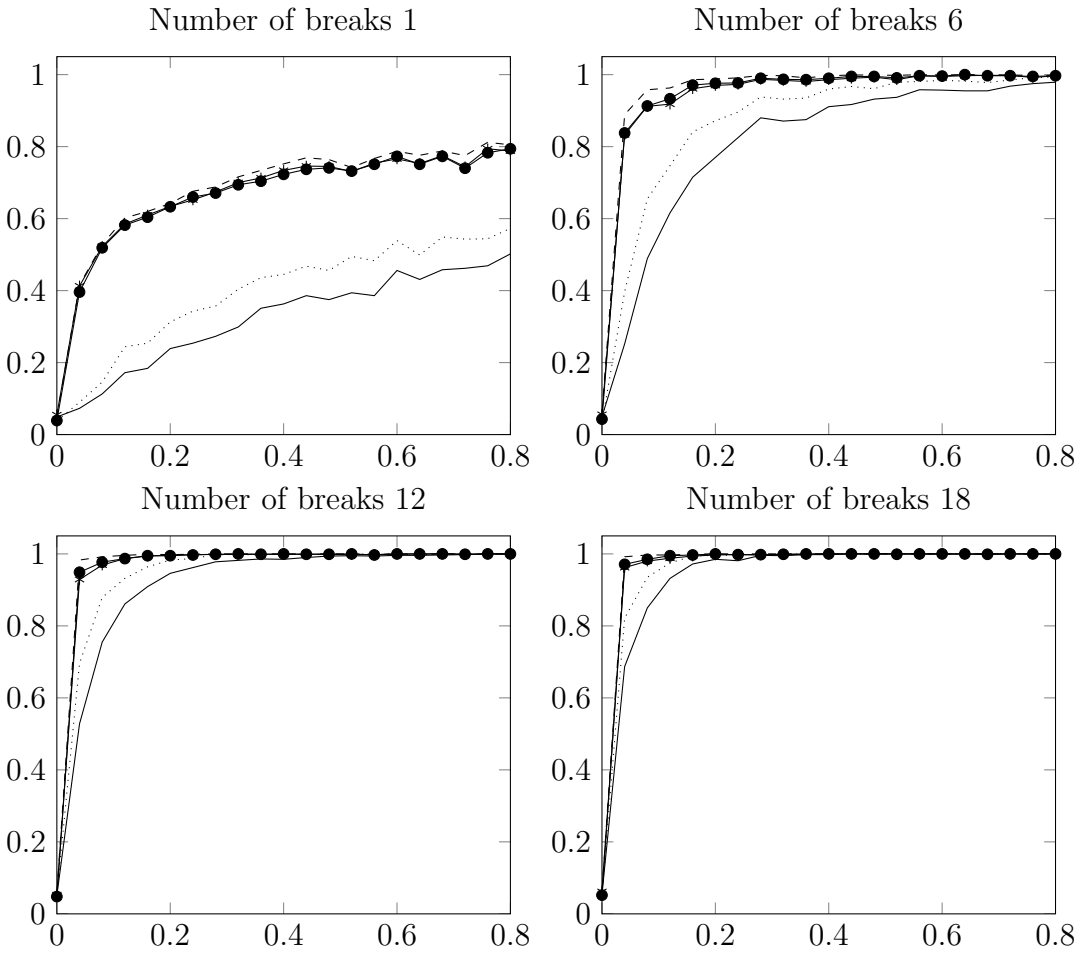

Figure 14: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for random breaks.

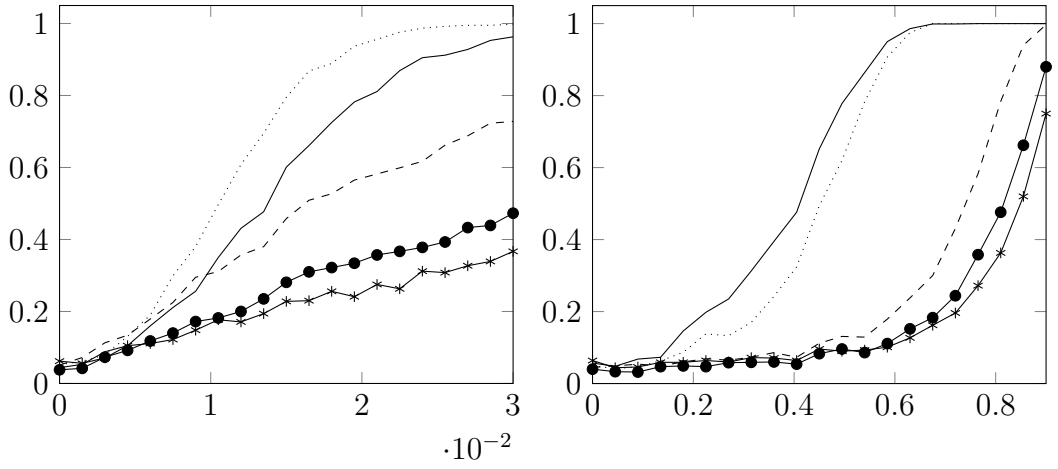

Figure 15: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for state space.

t–distribution time–varying variance

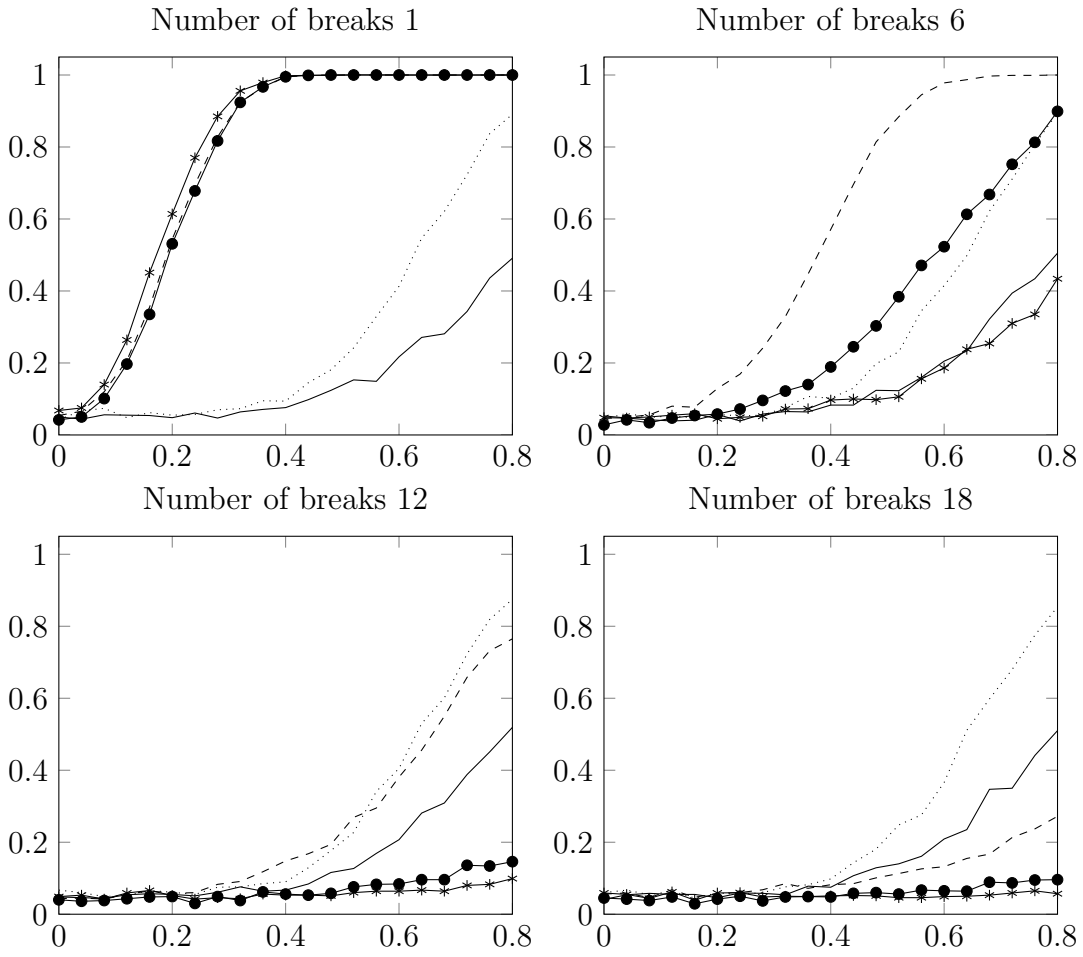

Figure 16: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for regime–switching.

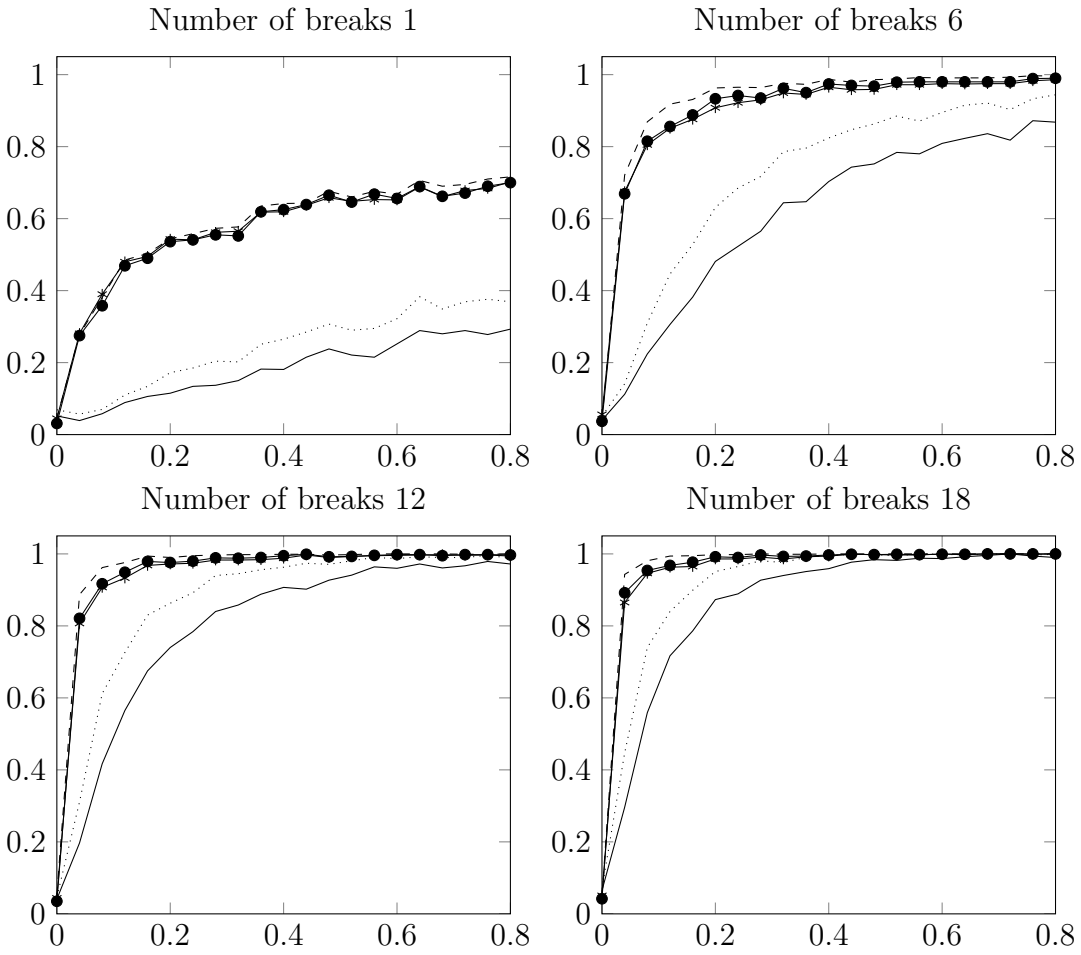

Figure 17: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for random breaks.

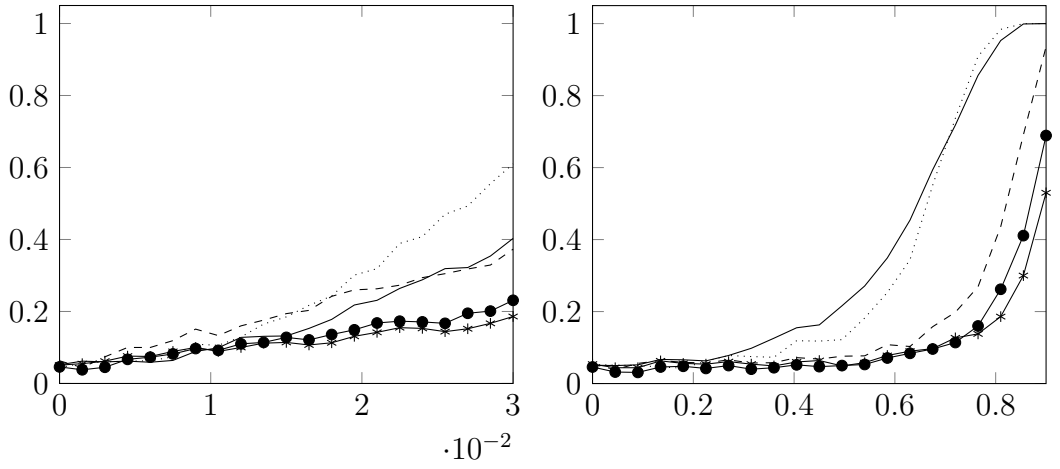

Figure 18: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for state space.

t–distribution time–varying correlation

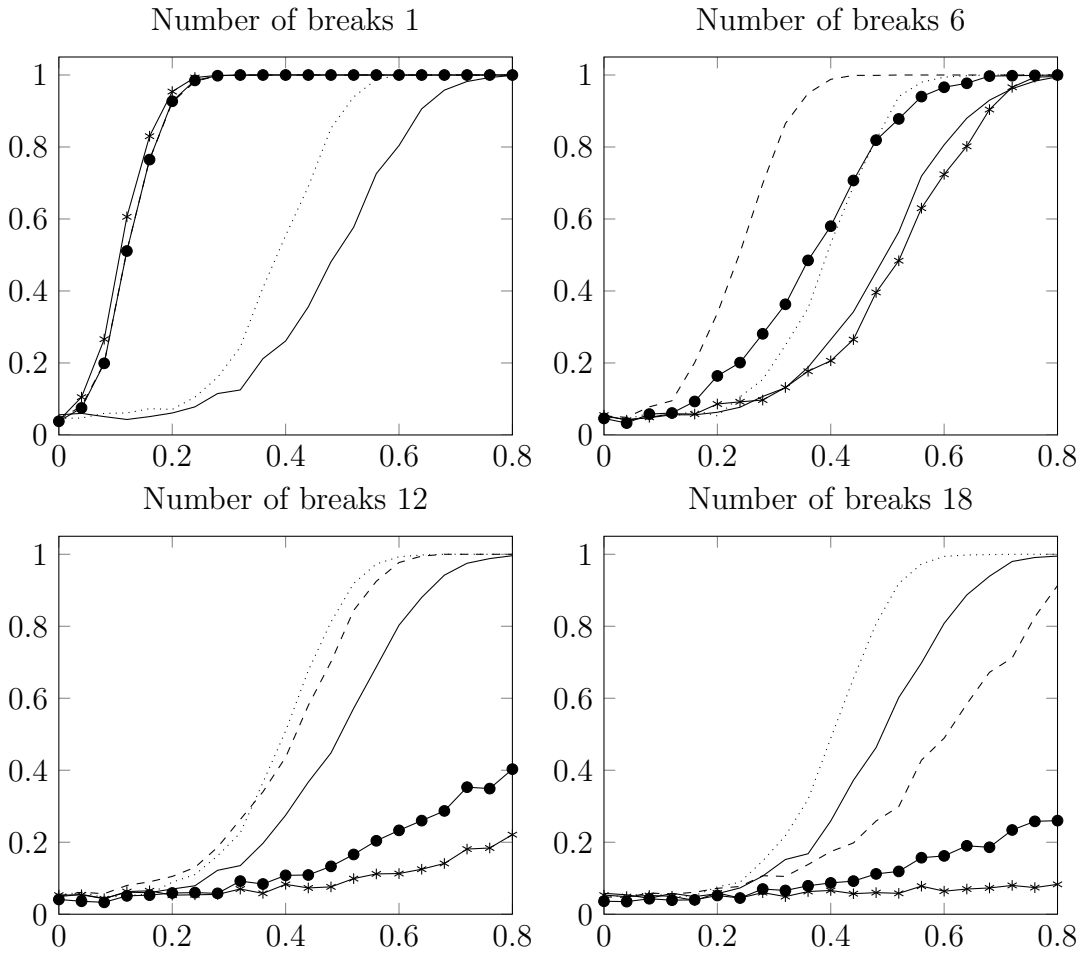

Figure 19: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for regime–switching.

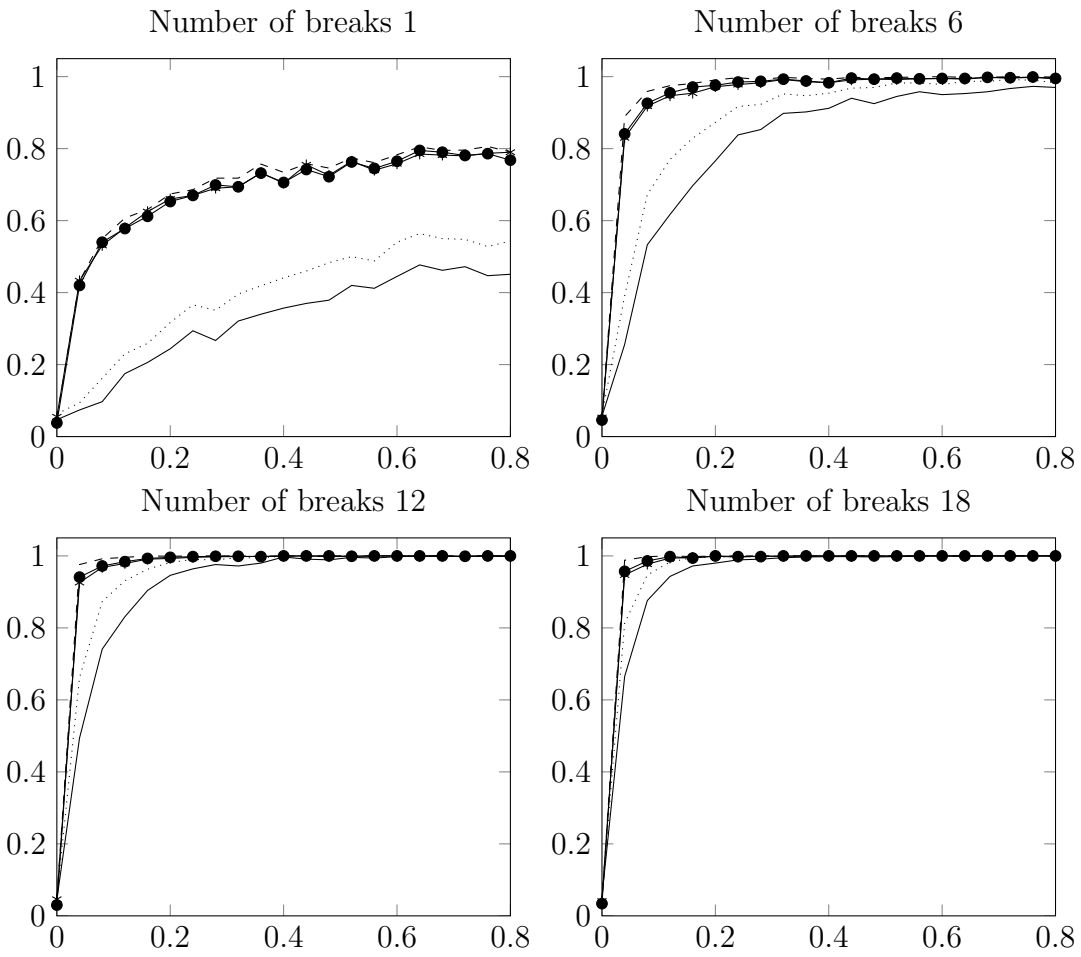

Figure 20: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for random breaks.

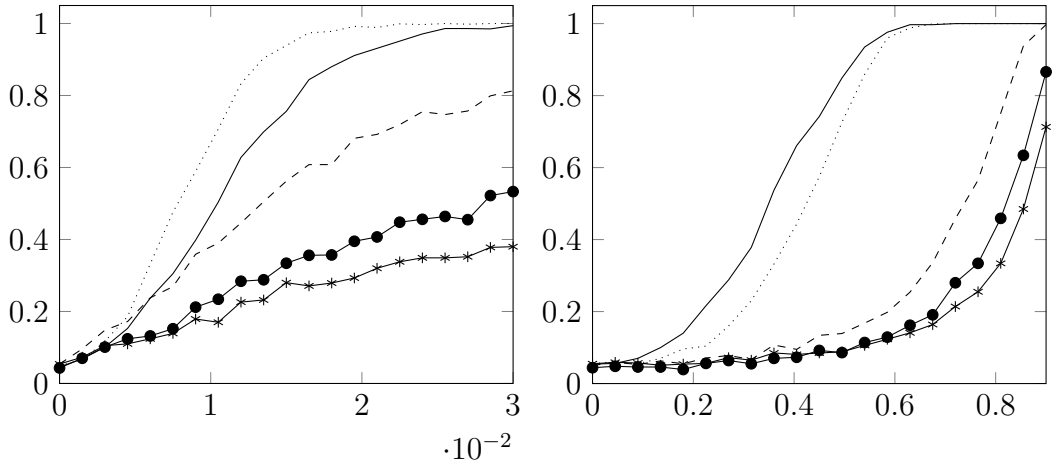

Figure 21: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for state space.

Beta first setting

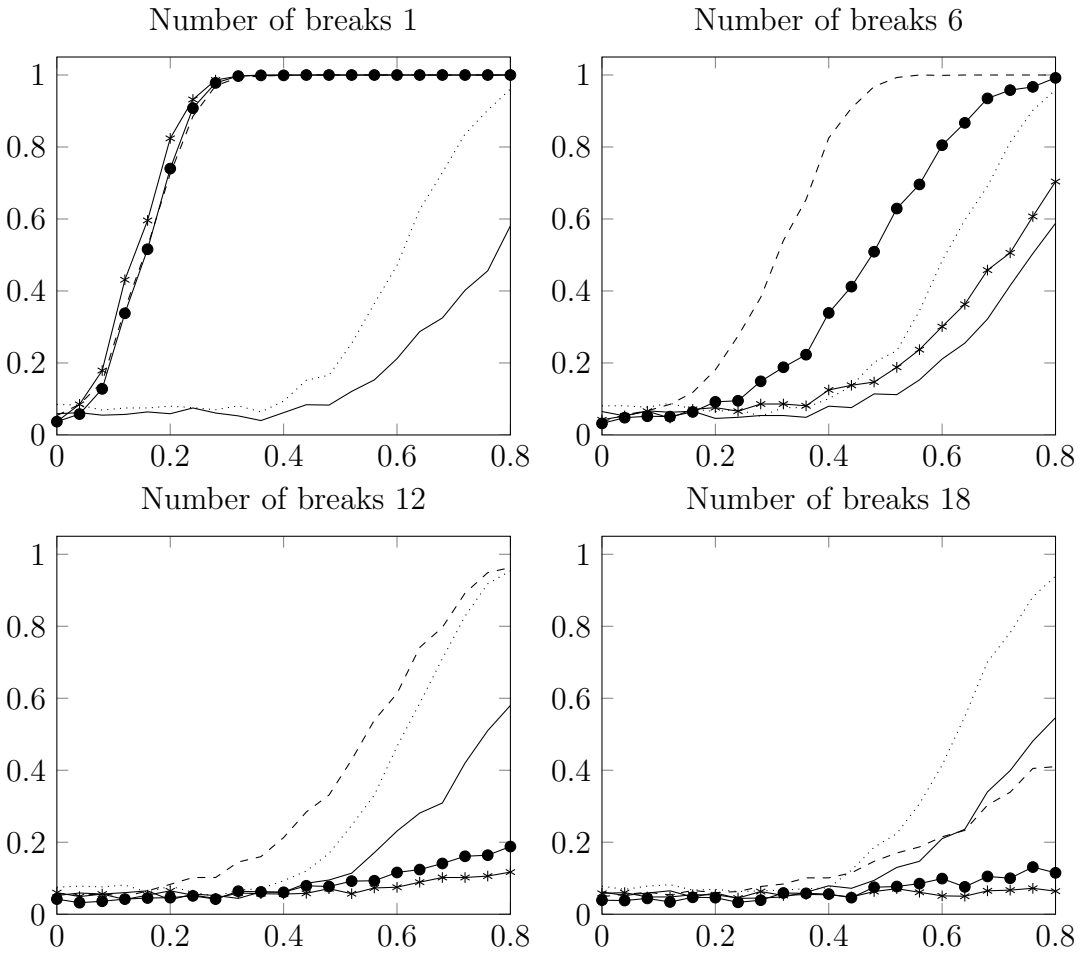

Figure 22: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for regime–switching.

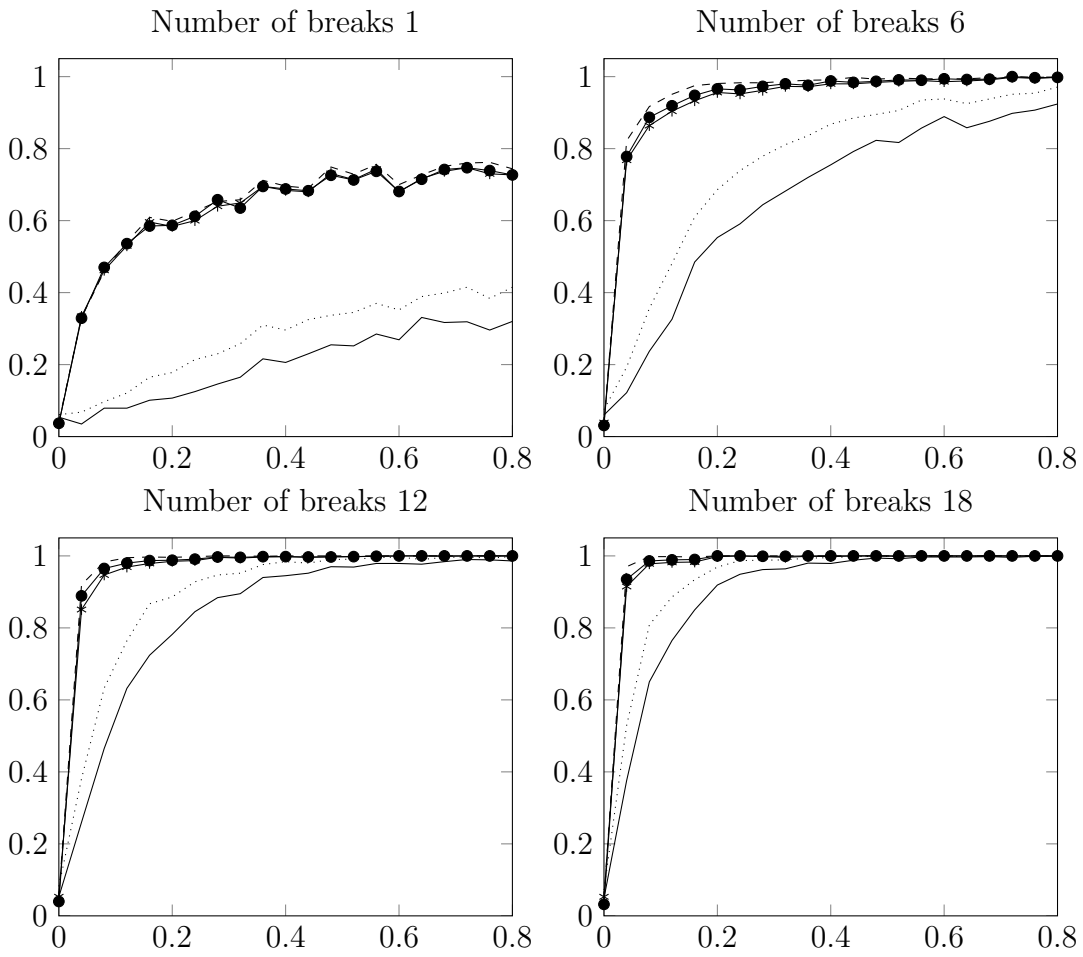

Figure 23: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for random breaks.

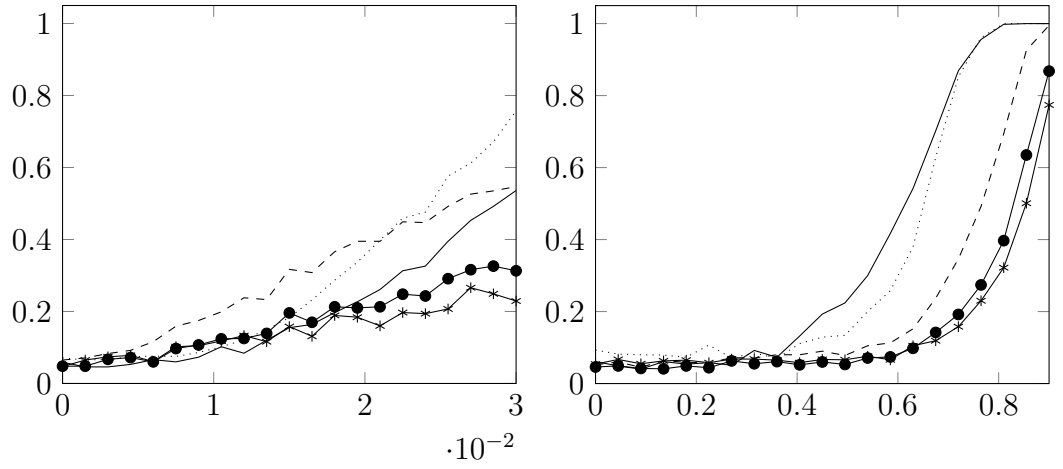

Figure 24: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for state space.

Beta second setting

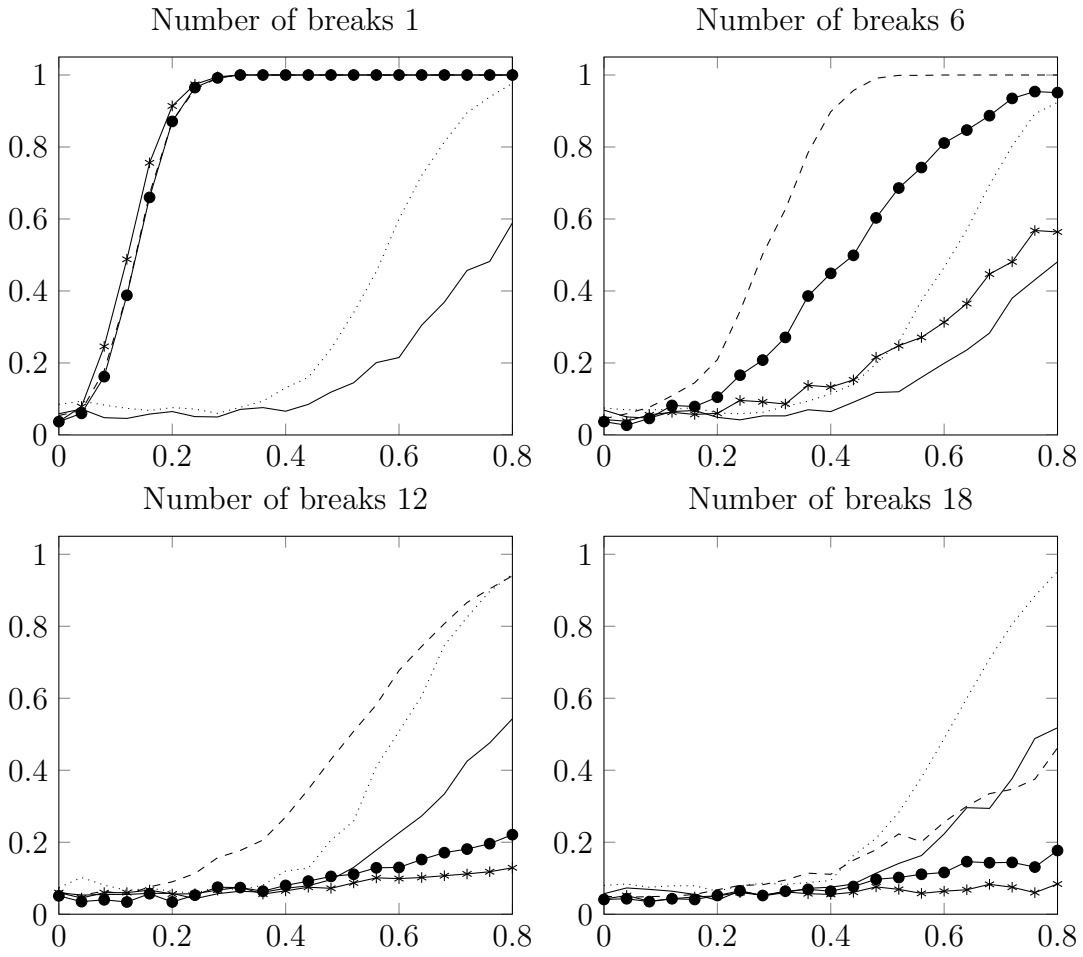

Figure 25: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for regime–switching.

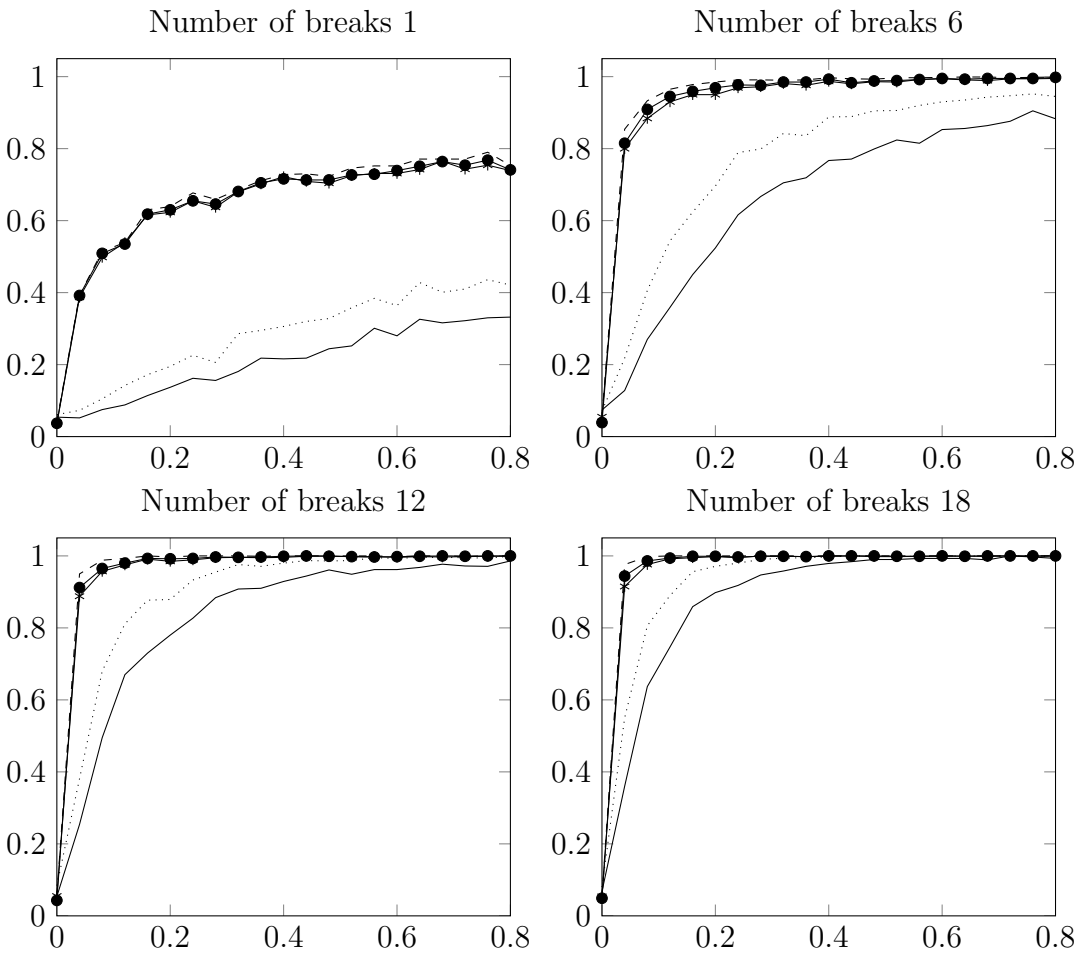

Figure 26: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for random breaks.

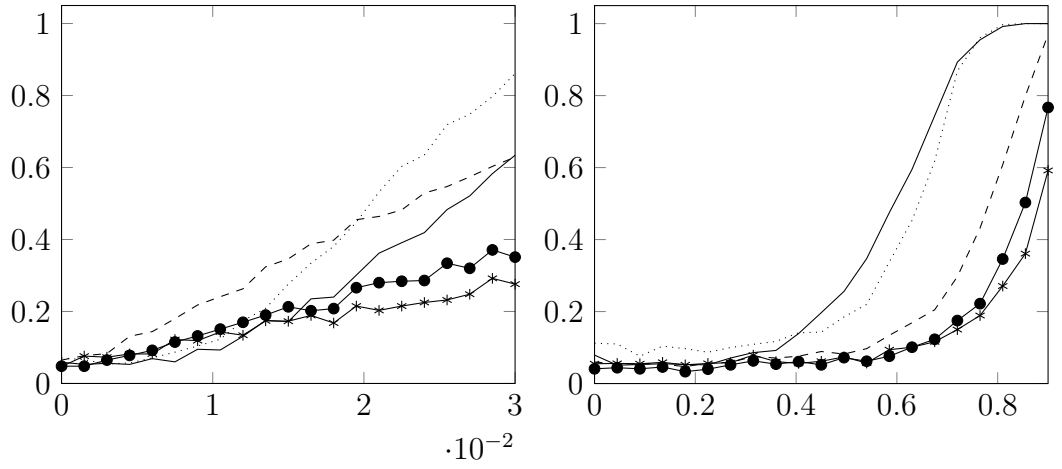

Figure 27: Empirical power functions of the $LM_{GAS(0,1)}$ (—), the $LM_{GAS(0,5)}$ (···), the Andrews test (•–), the Müller-Petalas test (- - -) and the Nyblom test (∗–) for state space.