Testing for Parameter Instability across Different Modeling Frameworks

Francesco Calvori\textsuperscript{a}, Drew Creal\textsuperscript{b}, Siem Jan Koopman\textsuperscript{c,d}, André Lucas\textsuperscript{c}

\textsuperscript{(a)} R&D, UnipolSai Assicurazioni, Bologna, Italy
\textsuperscript{(b)} Booth School of Business, University of Chicago, USA
\textsuperscript{(c)} VU University Amsterdam and Tinbergen Institute, The Netherlands
\textsuperscript{(d)} CREATES, Aarhus University, Denmark

Abstract

We develop a new parameter instability test against the alternative of a time-varying parameter. The new test generalizes the seminal ARCH-LM test for a constant variance against the alternative of autoregressive conditional heteroskedasticity to settings with nonlinear time-varying parameters and non-Gaussian distributions. These alternative model settings are unified by the class of score-driven time-varying parameter models. We discuss the performance of our test compared to classic but also to more recent alternative tests, including tests against structural breaks and parameter-driven dynamics. For regime switching and mean-reverting parameter-driven dynamics, the new test has higher and more stable power properties. For more local alternatives and for structural breaks, the recent test by Müller and Petalas (2010) performs best. We provide an application to a heavily unbalanced panel of losses given default for U.S. corporations from 1982 to 2010 and provide evidence of significant parameter instability in the parameters of a static beta distributed model.

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

JEL classifications: C12, C52, C22.

\*We thank Bob Brugman for his excellent research assistance. Furthermore, we have benefited from the comments of participants at the Workshop on Score Driven Models in Amsterdam. D. Creal thanks the William Ladany Faculty Scholar Fund at Booth School of Business for financial support. S. J. Koopman acknowledges support from CREATES, Center for Research in Econometric Analysis of Time Series (DNRF78), funded by the Danish National Research Foundation. A. Lucas thanks the Dutch National Science Foundation (NWO Grant VICI453-09-005) 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 the empirical modeling of financial time series is parameter instability. Hansen (2001) provides an overview of a large number of different parameter instability tests, 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 instability, we can consider different model specifications under the alternative. For example, there might be one or more deterministic structural breaks in the parameters as in, for example, Vogelsang and Perron (1998), Bai and Perron (2003), Perron (2006), and Qu and Perron (2007); the parameters may exhibit regular regime switches as in Hamilton (1989); or the parameters may evolve continuously over time, either in a parameter-driven (state space) framework such as Harvey (1989), Bauwens and Veredas (2004), Shephard (2005), Hafner and Manner (2012), and 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), Creal et al. (2013), and Harvey (2013).

The goal of this paper is twofold. First, we develop a new test for parameter instability in nonlinear and non-Gaussian models against the alternative of a generalized autoregressive score (GAS) model of Creal et al. (2013) and Harvey (2013). GAS models provide a flexible framework for the introduction of time-varying parameters in a model and include the generalized autoregressive conditional heteroskedasticity (GARCH) model of Engle (1982) and Bollerslev (1986), the autoregressive conditional duration (ACD) model of Engle and Russell (1998), and the dynamic Poisson model of Davis et al. (2003) as special cases. The GAS model is a score-driven model that is characterized by a parametric conditional observation density. Since the GAS model is observation-driven, its likelihood function is available in closed-form, such that parameter estimation and inference are straightforward. A variety of empirical studies have illustrated that the GAS framework can successfully capture time-variation in key parameters; examples are given by Creal et al. (2011) and Lucas et al. (2014) for multivariate volatility and correlation models, Creal et al. (2014) for mixed measurement factor models, Harvey and Luati (2014) for location and scale models, and De Lira Salvatierra and Patton (2015) for copula models.\footnote{A more extensive compilation of papers on autoregressive score-driven models is presented at http://www.gasmodel.com.} GAS models have a good out-of-sample forecasting accuracy compared to nonlinear non-Gaussian state space models, even in cases where the latter are correctly specified; see Koopman et al. (2015). Moreover, GAS models have information theoretic optimality properties; see Blasques et al. (2015). A parameter instability 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.

Our new Lagrange Multiplier (LM) test is highly intuitive and practical. It tests for non-zero serial dependence in the score function of $\ell_t$ where $\ell_t$ is the $t$th contribution to the loglikelihood function $\sum_{t=1}^{T} \ell_t$ in a static model for a time series of length $T$. We view our test as an omnibus diagnostic tool and as a generalization of the familiar ARCH-LM test of Engle (1982) to settings beyond time-varying volatility. The asymptotic distribution of the test is standard and follows by familiar results from White (1987). Similar to most omnibus LM diagnostic tests, our test can easily be computed by means of an auxiliary regression.

The second goal of our paper is to investigate the finite sample properties of the new test relative to other recent (and less recent) tests from the literature. We consider linear as well as nonlinear models where computational efficiency becomes a concern. The estimation of parameters in an extensively nonlinear model is typically cumbersome and computationally demanding. Hence we favor the use of LM-based test statistics. In particular, we consider the recently proposed test against local parameter-driven time variation of Müller and Petalas (2010). As further benchmarks, we include the sup-LM test of Andrews (1993), and the test against random walk parameter alternatives of Nyblom (1989). All these tests are applicable to both linear and nonlinear settings. However, each of these tests has a different specification for the time-varying parameter under the alternative hypothesis.

The results of Müller and Petalas (2010) are of particular interest to our new test. These authors prove that a test against a parameter-driven alternative is asymptotically optimal against very general local alternatives, whether parameter-driven or not. 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 time-varying parameters, and stationary parameter-driven dynamics. Second, we pay attention to non-local alternatives. The theoretical results in Müller and Petalas (2010) do not provide much insight into the behavior of alternative testing methodologies under non-local alternatives in finite samples. In particular, the finite sample performance of the different testing procedures may crucially depend on the type of data generating process under the alternative. Finally, we also consider versions of their test based on alternative values of their tuning parameters and show that this may further increase the power of their test.

We obtain several interesting insights. The new test against the observation-driven GAS alternative exhibits higher power for alternatives with frequent regime switches or non-local parameter-driven time-variation. In cases where the parameter-driven behavior is close to the
null or where infrequent structural breaks occur, the test of Müller and Petalas (2010) performs best. The optimality of the Müller-Petalas test for parameter-driven time variation close to the null follows immediately from the analytical results in Müller and Petalas (2010). The good performance in case of infrequent breaks is consistent with their simulation results that compare their test to the sup-\textit{LM} test of Andrews (1993). When the size or frequency of time-variation is higher, we show that the new GAS-LM test against the observation-driven alternatives starts to perform better. The power performance of our test is also surprisingly robust over alternative specifications of the data generating process; this stands in contrast to most of the GAS test’s competitors.

The results can be explained by looking into the construction of the different test statistics. The GAS-LM test relies directly on the serial dependence in the scores of the loglikelihood from the static model. These serial correlations can typically be estimated accurately even when the data generating process of the time-varying parameter is erratic and moves quickly. Hence we find a good performance of our test in settings with many breaks or with 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 the assumption of a highly persistent parameter process. When the true time-varying parameter moves quickly, the estimated path typically becomes almost constant, thus reducing the estimated volatility of the path and the power of the test. This phenomenon of a bad estimate of the path of the time-varying parameter also applies to the GAS model, but it does not affect the GAS-LM test since the latter is not based on the volatility of the estimated path of the time-varying parameter. Conversely, if the true path of the parameter evolves slowly over time, the likelihood ratio nature of the Müller and Petalas (2010) test leads to a better performance compared to that of the GAS-LM test.

In our empirical illustration, we test for the instability of parameters in loss given default (LGD) models that are used for credit risk analyses. Loss given default is the fraction of the outstanding amount of a loan or bond that is lost if the company defaults. It is a key ingredient of current models for 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 distribution for modeling LGD fractions. The use of static parameters in this application may be inappropriate since financial conditions vary over time and affect the LGDs accordingly. For example, losses may be higher on average in situations where default risk is also higher, thus exacerbating the total loss experience. If the parameters of a model for LGDs are actually time-varying, the regulator may require higher capital requirements for financial institutions to mitigate financial stability concerns.
We analyse a panel data set of LGDs for corporate bond data obtained from Moody’s to test for the presence of time-varying parameters. The data set is non-standard and provides an interesting example of the flexibility of our testing approach. The number of LGD observations for each quarter varies over time because LGDs can only be observed when a default occurs, and the number of defaults evidently varies over time. Assuming that the LGDs are drawn from a beta distribution with possibly time-varying parameters, all tests strongly confirm that the distributional properties of LGDs vary over time. In particular, we find that LGDs have been on average very low compared to the static model during the period leading up to the 2008 financial crisis. It suggests that the abundance of liquidity during this period has not only prevented firms from defaulting, but has also mitigated the losses for those cases in which a default was unavoidable.

The remainder of this paper is organized 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 provide simulation results. Section 4 discusses the empirical application. Section 5 concludes. The online Appendix presents additional simulation results.

2 Testing frameworks for time-varying parameters

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. The dimensions $m$, $k$ and $n$ are positive integers; for a univariate model with a single time-varying parameter and four static parameters, we have $m = k = 1$ and $n = 4$.

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. It leads to estimation and inference procedures that can be easily implemented. The main challenge however is to determine the function of the observations that drives the parameter $f_t$ through time. A general approach encompassing many popular nonlinear and non-Gaussian dynamic models is the generalized autoregressive score (GAS) model of Creal et al. (2013); see also Harvey (2013). In the GAS($p, q$) model, the
observations $y_t$ have the dynamic specification

$$y_t \sim p(y_t|f_t; \delta)$$

(1)

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

(2)

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

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

(3)

where $\nabla_{f,t}$ is the score of the conditional observation density. The $k \times k$ matrix $S_t = S(f_t; \delta)$ is the scaling matrix for the score. For example, we can consider 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.

The defining feature of the GAS model is its use of the score of the conditional observation density to drive the parameter $f_t$ through time. At each time period, the dynamics of the time-varying parameter can be interpreted as a steepest-ascent or Gauss-Newton step, where the local fit of the model is improved by using the information in the most recent observation and its distribution. Such steps locally improve the Kullback-Leibler divergence of the model even in cases where the model is mis-specified; see Blasques et al. (2015). The GAS framework encompasses as special cases the 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. (2013), the models for Poisson counts of Davis et al. (2003), and the Beta-t-GARCH model of Harvey (2013), among many others.

To introduce 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

$$y_t = x_t^\prime \beta + \sigma_t \varepsilon_t$$

where $\varepsilon_t$ has mean zero and variance one, tests the null of a constant variance against the alternative

$$\sigma_t^2 = (1 - A) \omega + A \varepsilon_t^2 = (1 - A) \omega + A(\varepsilon_t^2 - \sigma_t^2) + A \sigma_t^2,$$

(4)

with parameter $|A| < 1$, for $t = 1, \ldots, T$. Under the alternative hypothesis, the variance $\sigma_t^2$ varies around the static level $\omega$ as driven by the scaled score $\varepsilon_t^2 - \sigma_t^2$ of a Gaussian density with respect to the parameter $\sigma_t^2$. The null hypothesis is $H_0 : A = 0$.

The LM test against a GAS alternative takes the same perspective as the ARCH-LM test (4), but generalizes the observation density and allows the time-varying parameter $f_t$ to characterize a different distributional property than the variance. Formally, we test the null
hypothesis of no parameter variation against the GAS alternative

\[ f_{t+1} = (1 - A_1 - \ldots - A_q)\omega + \sum_{i=1}^{q} A_is_{t-i+1} + \sum_{i=1}^{q} A_if_{t-i+1}, \]

where the dynamics of \( f_t \) are driven by the scaled scores \( s_t \) from (3). Similar to (4), under the alternative the time-varying parameter \( f_t \) varies around its static level \( \omega \). We can use the same arguments as in Lee (1991) to allow for different coefficients \( B_i \) (rather than \( A_i \) only) for the lags of \( f_{t-i+1} \) under the alternative, for \( i = 1, \ldots, q \).

To define the LM test statistic, let \( \ell_t(\delta, \omega, a) = \ln p(y_t|f_t; \delta) \) be the loglikelihood contribution at time \( t \), where we suppress the dependence of \( f_t \) on the static parameters \( \delta, \omega \), and \( a = \text{vec}(A_1, \ldots, A_q) \). Define \( \bar{s}_{p,t} = \text{vec}(s_t, \ldots, s_{t-p+1}) \) and let \( G'_t = (\nabla'_{\delta,t}, \nabla'_{\omega,t}, \nabla'_{\omega,t} \otimes \bar{s}'_{p,t-1}) \) where \( \otimes \) is the Kronecker product and with \( \nabla_{\delta,t} \) and \( \nabla_{\omega,t} \) denoting the derivatives of \( \ell_t \) with respect to \( \delta \) and \( \omega \), respectively. Following White (1987), the LM test for \( H_0 : a = 0 \) versus the alternative \( H_1 : a \neq 0 \), is given by

\[ \text{LM} = G'H^{-1}G, \quad G = \sum_{t=1}^{T} G_t, \quad H = \sum_{t=1}^{T} G_t G'_t, \]

where derivatives are evaluated at the maximum likelihood estimates under the null hypothesis. The covariance matrix \( H \) can be replaced by a robust long-run covariance matrix, that is

\[ \hat{H} = \sum_{t=1}^{T} \sum_{\tau=1}^{t} w_{T,t-\tau} (G_t G'_\tau + G'_t G_t), \]

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

Under standard regularity conditions, the GAS-LM test converges under the null to a \( \chi^2 \) distributed random variable with \( \text{dim}(a) \) degrees of freedom; see White (1987). An appealing feature of the new test is thus that its asymptotic statistical theory is entirely standard. This contrasts with the asymptotic behavior of some of the other tests that are considered later.

Following Davidson and MacKinnon (1990), the LM test statistic can be written as the explained sum of squares of the auxiliary least squares regression

\[ 1 = G'_t\beta_{LM} + \text{residual} \]

\[ = \nabla'_{\delta,t}\beta_{LM}^{\delta} + \nabla'_{\omega,t}\beta_{LM}^{\omega} + (\nabla_{\omega,t} \otimes \bar{s}_{p,t-1})' \beta_{LM}^{a} + \text{residual}, \]

where \( \beta_{LM} = (\beta_{LM}^{\delta}, \beta_{LM}^{\omega}, \beta_{LM}^{a}) \) is a vector of regression parameters and all derivatives in \( G_t \) are evaluated under the null. The regression interpretation of the GAS test makes it easy to compute in standard statistical software. The first derivatives of the conditional observation density at each time \( t \) can be obtained either analytically or numerically.

7
The GAS-LM test has an intuitive interpretation. The key term on the right-hand side of the auxiliary regression (7) is \( \nabla_{\omega,t} \otimes \bar{s}_{t-1} \). The elements of this vector are vec\((S_{t-i} \nabla_{\omega,t-i} \nabla'_{\omega,t})\), for \( i = 1, \ldots, q \), because, under the null, the score of the conditional density with respect to \( f_t \) is equal to the score with respect to \( \omega \). Hence, the LM test against the GAS alternative verifies whether there is any serial dependence in the scores \( \nabla_{f,t} \) of the static model. If serial dependence exists, the actual autocorrelations in the likelihood scores can be exploited to improve the fit of the model by using the scores to drive the time-varying parameter \( f_t \). This is exactly what the dynamics of the GAS model in (2) achieve.

Although the above LM test has been derived with the GAS alternative in mind, we expect it to have power against other forms of parameter instability as well. The test can therefore be regarded as a generic portmanteau test against model misspecification. The same holds for tests against structural break alternatives and against parameter-driven time variation. Next we discuss these more general applications of our test.

### 2.2 Parameter-driven time-variation

For parameter-driven models, the time-varying parameter \( f_t \) is a stochastic process that is subject to its own source of error. Important examples of this class of models are unobserved components time series models as in Harvey (1989), stochastic volatility models as reviewed in Shephard (2005), stochastic conditional duration models as in Bauwens and Veredas (2004), and stochastic copula models as in Hafner and Manner (2012) and Creal and Tsay (2015). The randomness in both \( f_t \) and \( y_t \) lead to challenging problems in parameter estimation and testing. The likelihood function is not available in closed form except in cases such as linear Gaussian state space models and discrete-state hidden Markov models; see Durbin and Koopman (2012) and Hamilton (1989), respectively. In all other cases likelihood-based inference requires approximation and/or simulation methods; see, for example, Creal (2012) and Durbin and Koopman (2012) for discussions on such methods.

Müller and Petalas (2010), denoted as MP10 hereafter, provide an elegant and general framework for testing parameter instability. Their approach encompasses nonlinear and non-Gaussian models with moderately time-varying parameters. If the time variation vanishes asymptotically at the appropriate rate, MP10 show that we can address the inference problem of parameter instability 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

The key intuition of the MP10 test comes from a (pseudo-)linear Gaussian state space model
with observation and transition equations given by

\[
H V^{-1} \nabla_{\omega,t} = S^{-1}(f_t - \bar{f}) + \nu_t, \quad \nu_t \sim N(0, S^{-1})
\]

\[
(f_{t+1} - \bar{f}) = (1 - cT^{-1})(f_t - \bar{f}) + \tilde{\nu}_t, \quad \tilde{\nu}_t \sim N(0, c^2 T^{-2} S^{-1}),
\]

where

\[
H = T^{-1} \sum_{t=1}^{T} \frac{\partial^2 \ln p(y_t|\bar{f}; \delta)}{\partial f_t^2}, \quad V = T^{-1} \sum_{t=1}^{T} \nabla_{\omega,t} \nabla_{\omega,t}', \quad S = H^{-1}VH^{-1}.
\]

The loglikelihood score \( \nabla_{\omega,t} \) is the same as for the observation driven model and the state
variable \( f_t \) follows a nearly-integrated process with fixed tuning parameter \( c \). The parameter \( \bar{f} \)
is a fixed benchmark level for \( f_t \), for \( t = 1, \ldots, T \). It follows that \( f_t \) is a persistent process with
local time variation. As the sample size grows, time variation in \( f_t \) vanishes as the autoregressive
parameter converges to unity and the variance of the transition equation in (8) converges to
zero.

The variable \( H V^{-1} \nabla_{\omega,t} \) can be viewed as a pseudo-observation. Its linear Gaussian state
space model (8) is the result of applying Laplace transformations to nonlinear and non-Gaussian
models. A similar technique is used for the likelihood-based approaches of Shephard and Pitt
(1997) and Durbin and Koopman (1997, 2000) where they adopt an approximating linear,
Gaussian model to implement importance sampling or MCMC methods. Such simulation-
based methods for the estimation of parameters in nonlinear non-Gaussian state space models
are used extensively in econometrics. MP10 point out that the key difference between the
and their approximating model is the use of the global Hessian \( H \) rather than the local
Hessian of the conditional observation density at time \( t \).

Müller and Petalas (2010) construct a point optimal (likelihood ratio) test of the null \( c = 0 \)
versus the alternative \( c = 10 \). Although the theory in MP10 is somewhat advanced, the
proposed test statistic is surprisingly straightforward to compute using regression methods. An
algorithm is provided in their paper. The point optimality of the test allows a likelihood ratio
test interpretation. As a result, the test has a power advantage compared to an \( LM \) 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. Consequently, the test captures part of the gain of the likelihood ratio compared to
the Lagrange multiplier test, just as if parameters of a model would have been estimated under both the null and the alternative rather than under the null only.

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

The crucial ingredient of the MP10 test is the sum of $(f_t - \bar{f}) \cdot \nabla_{\omega_t}$. The test uses the variability of the estimated path $f_t$ around its static counterpart $\bar{f}$ for $c = 10$. The differences $(f_t - \bar{f})$ are weighted by the likelihood score with respect to the, possibly, time-varying 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. The smoothed estimate of the path for the test is obtained under $c = 10$, which implies a high degree of persistence for sufficiently large $T$. 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 changes. In these cases, the estimated path of $f_t$ can become close to a constant, resulting in a low value of the test statistic and a low power of the corresponding testing procedure.

Compared to the GAS-LM test of 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 weighs 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 tuning parameter $c = 10$. Third, the distributions of the GAS and MP10 tests under the null differ profoundly. The GAS-LM test follows the standard $\chi^2$ asymptotics of White (1987) while the MP10 test follows the asymptotic distribution as derived in Elliott and Müller (2006).

### 2.3 Structural breaks

Andrews (1993) proposes a general parameter instability test for nonlinear parametric models against alternatives with a one-time break in (a subset of) the parameters. Generalizations to multiple breaks are possible, but are typically computer intensive unless the structure of the model is sufficiently simple; see Bai and Perron (2003). The tests against a structural change alternative are based on partial-sample GMM estimators and can be of the supremum Wald, Lagrange multiplier ($LM$), and likelihood ratio types. Modifications of these tests that use weighted averages rather than the supremum of the tests over all possible break points are
proposed by Ploeger et al. (1989) and Andrews and Ploeger (1994). Here we focus on the $LM$ based version of the test. This precludes the need to estimate a possibly nonlinear model over many different subsamples. It can lead to a time-consuming process with many computations, in particular 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 hypotheses for the Andrews’ sup-$LM$ test are given by

\[
H_0 : f_t = \bar{f}_0 \quad \forall t \geq 1 \text{ and some } \bar{f}_0 \in F \subset \mathbb{R}^k, \tag{9}
\]

\[
H_1 : \bigcup_{\pi \in \Pi} H_{1,T}(\pi) \text{ for some } \Pi \subset (0, 1), \tag{10}
\]

\[
H_{1,T}(\pi) : f_t = \begin{cases} 
\bar{f}_1(\pi), & \text{for } t = 1, \ldots, \lfloor \pi T \rfloor, \\
\bar{f}_2(\pi), & \text{for } t = \lfloor \pi T \rfloor + 1, \ldots, T,
\end{cases} \tag{11}
\]

for constants $\bar{f}_1(\pi), \bar{f}_2(\pi) \in F$. The test is designed for a single break at an unknown date. However, the test also has good power properties against a range of more general alternatives; see, for example, the survey of Hansen (2001). The distribution of the Andrews’ sup-$LM$ test is 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’ sup-$LM$ test does not build on the autocorrelations 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 alternative, which is a structural break at an unknown point in time. When regular switches between alternative values of the parameter occur, the Andrews’ test may have difficulty in identifying such a pattern. The sample means of moment conditions before and after any particular tentative breakpoint may fail to be sufficiently different in small sample sizes. We expect to observe this problem when the the true time-varying parameters are subject to regular regime switches or to strongly mean reverting dynamics.

### 2.4 Martingale type time–variation

Our final benchmark is the parameter instability test of Nyblom (1989). This “all-purpose” test is based on the assumption that under the alternative the time-varying parameter follows a martingale process. Nyblom (1989) argues that his test encompasses the case of one or more structural breaks. The Nyblom test is therefore related to both $LM$ tests of Sections 2.2 and 2.3.
The key ingredient of the Nyblom test is the partial sum of the likelihood scores. Hence the test is close to the partial sums in the Andrews’ test. However, the Nyblom test does not take a supremum, it relies on the average of the squares of partial sums. We therefore expect that the Nyblom test has an inferior performance compared to the other three tests in most settings.

3 Monte Carlo study

3.1 Design of study

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), and the test of Nyblom (1989). Müller and Petalas advise researchers to use the value $c = 10$ to compute their test. We call this the MP10 test. To check the sensitivity of the procedure to this choice, we also compute the test for $c = 5$ and $c = 20$, denoted as MP5 and MP20, respectively. We also use the procedure of Escanciano and Lobato (2009) to select an optimal number of lags in the GAS-LM test, denoted as GAS-LM($\ast$). The maximum lag length considered in GAS-LM($\ast$) is 20. We use $N = 10,000$ replications to build simulated power curves.

We use a wide range of alternatives. We differentiate the DGPs considered along two dimensions. First, we consider DGPs with different types of dynamics for the time-varying parameter. In particular, we have regime switching models, models with random structural breaks, and state space models. Second, we consider different degrees of time-variations in the model. In particular we consider time-variation in the mean, the variance, the dependence structure, and higher order moments of the observation density.

3.2 Data generation processes for parameters

We consider the following DGPs for the 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

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

with $\Delta$ denoting the difference (in absolute value) between the two regimes. For example, for $n_b = 4$, we have $f_t = \Delta$ for $[0.2T] + 1, \ldots, [0.4T]$, and for $[0.6T] + 1, \ldots, [0.8T]$, and zero
elsewhere. This creates regular and equally sized patches where the parameter alternately takes
the value 0 and $\Delta$. 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 $[\pi_j T] + 1$ for $j = 1, \ldots, n_b$,

$$f_t = \sum_{j=1}^{n_b} 1_{\{t > [\pi_j T]\}} v_j,$$

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

**State space models:** For a DGP with parameter-driven dynamics, we assume that $f_t$ follows
an autoregressive process of order one

$$f_{t+1} = \phi f_t + \sigma_\eta \eta_t,$$

where $\eta_t$ is normally distributed with zero mean and unit variance.

### 3.3 Data generation processes for observations

For each of the three different dynamic frameworks for $f_t$ in Section 3.2, we consider the
following observation models.

**Time-varying mean:** $y_t = f_t + \sigma_\varepsilon \varepsilon_t$, with disturbance $\varepsilon_t \in \mathbb{R}$ and standard deviation $\sigma_\varepsilon \in \mathbb{R}^+$;

**Time-varying log-variance:** $y_t = \exp(f_t/2) \varepsilon_t$, with disturbance $\varepsilon_t \in \mathbb{R}$;

**Time-varying dependence:**

$$y_t = \left( \begin{array}{cc} 1 & \tanh(f_t) \\ \tanh(f_t) & 1 \end{array} \right)^{1/2} \varepsilon_t,$$

with vector disturbance $\varepsilon_t \in \mathbb{R}^2$ and with the independent elements coming from either
the standardized normal density or the standardized Student’s $t(5)$ density;

**Time-varying beta distribution:** $y_t \sim \text{Beta}(\alpha_t, \beta_t)$, with both coefficients $\alpha_t$ and $\beta_t$ time-varying.
We also use the beta model in the empirical application of Section 4 where it is more thoroughly discussed. In the empirical application, both coefficients evolve independently over time using a two-dimensional $f_t$. In our current simulation setting we consider a simplified case, where both time-varying parameters in the beta model depend on a single time-varying scalar $f_t$. We consider two versions of this model. In the first version, we set

$$\alpha_t = \bar{f} \times \frac{\exp(f_t)}{1 + \exp(f_t)}, \quad \beta_t = \frac{\bar{f}}{1 + \exp(f_t)}, \quad \bar{f} > 0,$$

such that 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 are both time-varying. 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

$$\alpha_t = \bar{f} \times \exp(f_t), \quad \beta_t = (1 - \bar{f}) \times \exp(f_t), \quad 0 < \bar{f} < 1,$$

such that the mean $\mu_t = \bar{f}$ is constant and only the variance $\bar{f}(1 - \bar{f})/(1 + \exp(f_t))$ varies over time.

In total we have 24 simulation experiments: 3 time-varying parameter specifications that are combined with 8 model specifications (mean, log-variance and dependence, for the normal and Student’s $t$ densities, and two beta models). For each of these 24 experiments, we implement all tests at the 5% significance level. Andrews’ sup-LM test is implemented over a grid of breakpoints $\Pi$ in (10). We use the test based on the boundary breakpoint values of 15% and 85% of the sample size. Given the sample size, we can evaluate all breakpoints in between, even for the nonlinear models.

### 3.4 Results

Figure 1 presents the results for a regime switching mean and normally distributed errors. The top left panel shows that if there is only one regime switch then the Nyblom, Andrews, and MP tests have the highest power. This can be expected, as these tests are optimal for this case. The GAS tests are less powerful and need roughly between 2.5 and 3 times more distant alternatives than the Andrews test to obtain maximum power.

An important feature of the GAS-LM test is that its performance is hardly affected if the number of regime switches increases. This stands in sharp contrast to the performance of the other tests. The Nyblom test quickly loses power if we increase the number of regime switches. The Andrews sup-LM test also loses power, but more slowly. For 6 regime switches and 2,000 observations, the Nyblom test has similar power to the GAS-LM(1) test, while the Andrews
Figure 1: Power results for regime switches in mean for conditionally normally distributed errors

Empirical rejection frequencies for the GAS-based LM tests (with 1, 5, or optimal (∗) lags according to Escanciano and Lobato (2009)), the the Müller-Petalas test (with local alternative parameter 5, 10 (prescribed), and 20), the Andrew’s sup-LM test, and the Nyblom test. Clockwise, the panels are for 1, 6, 18, and 12 breaks.

test behaves similarly to the GAS-LM(5) test. The power of the MP tests also decreases, but still prevails for 6 regime switches. If the number of regime switches increases further, the performance of the MP tests also breaks, whereas that of the GAS-LM tests remains unaffected. For 18 switches, the MP10 test prescribed by MP10 has a worse power performance than the GAS-LM(1) test. Only the (non-prescribed) MP20 test in that case is still at par with the GAS-LM test. Further increases in the number of switches, however, will further deteriorate the performance of MP20.

The robustness of the GAS-LM test may seem surprising at first sight, but can easily be explained. The GAS test is based on the autocorrelation in the likelihood scores with respect to the time-varying parameter. The scores to compute this autocorrelation are evaluated under the null hypothesis. From the inspection of Figure 2, we learn that the parameter estimate under the null (the dotted straight lines) provides some average level of the true parameter path (the pulse function). As a result, the scores under the null roughly follow the pattern of the (demeaned) pulse function. This pattern 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 stable if we increase the number of switches.

The behavior is very different for the MP tests. For these tests, the estimated difference \((f_t - \bar{f})\) under the alternative \(c = 5, 10, 20\) plays a key role; see Section 2.2. Figure 2 plots the estimated path of \(f_t\) from MP10. An 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. When 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 3. For one break, the tests do not appear to reach a maximum power of one because the generated break dates are sometimes close to the starting or end point of the sample. The tests have little power against these alternatives. If the number of breaks increases, maximum power is reached quickly. This
Empirical rejection frequencies for the GAS-based LM tests (with 1, 5, or optimal (∗) lags according to Escanciano and Lobato (2009)), the Müller-Petalas test (with local alternative parameter 5, 10 (prescribed), and 20), the Andrew’s sup-LM test, and the Nyblom test. Clockwise, the panels are for 1, 6, 18, and 12 breaks.

may be due to the substantial probability that two consecutive breaks are in the same direction, which increases 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 compared to the GAS tests. This is in line with the objective of these tests: they are designed to be optimal against structural breaks.

Figure 4 presents the results for parameter-driven time-varying parameters as modeled by \( f_{t+1} = \phi f_t + \sigma_\eta \eta_t \), where \( \eta_t \sim N(0, 1) \). The left hand panel of Figure 4 contains the results for different values of the variance \( \sigma_\eta^2 \) on the horizontal axis and a fixed value of \( \phi = 0.9 \). The right hand panel contains the results for alternative values of \( \phi \) and a fixed value of \( \sigma_\eta^2 = 0.15 \).

In both settings, the GAS tests display the best overall power performance. For the case of fixed \( \phi = 0.9 \) (left panel), the true simulated parameter path exhibits strong mean reversion. These results are similar to the regular regime switches in Figure 1. We have already indicated why the MP tests have worse power performance compared to the GAS tests in this setting. The same phenomenon applies to the left panel in Figure 4 where for local alternatives, that is \( \sigma_\eta^2 \) being close to zero, the MP10 and MP20 tests have a better power performance than the GAS.
Figure 4: Empirical power functions for the Gaussian time-varying mean model with parameter-driven time variation \( f_{t+1} = \phi f_t + \sigma_\eta \eta_t, \eta_t \sim N(0, 1) \), and \((\sigma_\eta^2 = 0.00 \ldots 0.03, \phi = 0.9)\) and \((\sigma_\eta^2 = 0.15, \phi = 0.00 \ldots 1.00)\) in the left-hand and right-hand panel, respectively.

tests. These findings in Figure 1 confirm the analytical results of Müller and Petalas (2010) who prove that their test is optimal against local alternatives. For more distant alternatives, however, we find that the power of the GAS test is superior.

The right panel of Figure 4 provides further evidence that the GAS tests have the best overall performance. For large values of \( \phi \) (on the horizontal axis), the time variation becomes a martingale process. The Nyblom, Andrews, and MP appear to be only able to distinguish strongly persistent processes. For lower values of \( \phi \), these tests hardly have any power. By contrast, the power of the GAS-LM test quickly identifies time-variation, also for small values of \( \phi \). The power comes from the fact that even for strong mean reversion the score under the null still displays significant autocorrelation. The lack of power for the other tests is explained by the fact that for small values of \( \phi \) there is no substantial change in the level of the score under the null (the Andrews and Nyblom tests), nor is there a strong time variation in \((\hat{f}_t - \bar{f})\) in this case; compare Figure 2.

The online Appendix to this paper contains all additional simulation results. The results are highly robust and in line with the findings presented above. For example, the same conclusions can be made if we replace the Gaussian distribution by a Student’s \( t(5) \) distribution. The results for nonlinear models, such as models for time-varying variances and correlations, lead to similar conclusions, as well as the results for the time-varying beta models.

4 Empirical application

We consider quarterly observations of loss given defaults (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 losses 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. The percentage price drop is also known as market implied LGD and can become negative, for example, after a timely restructuring of a firm or after a merger announcement. We censor negative observed LGDs to 1 basis point, 0.01%. The censoring affects only 14 out of 1125 observations, that is 1.25% of our data set.

Our aim is to test whether there is significant time variation in the distributional characteristics of LGDs. The study into possible time-varying behavior of LGDs is important for credit risk modeling and financial stability research: 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 displays several non-standard features. First, LGDs are measured as percentage losses so that they are bounded to the interval $[0, 1]$. We therefore assume that the LGDs are drawn from a beta distribution. Second, the number of observed LGDs varies per quarter. Hence the dimension of the observation vector is typically different for each quarter. Such 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 maximum of 58 in 2009.

Let $y_{i,t}$ denote the $i$th 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 time-varying 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

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

where $\Gamma$ denotes the gamma function. The conditional score and information matrix for (15) are given by

$$
\nabla_{f,t} = \sum_{i=1}^{K_t} \left( \begin{array}{c} (\Psi(\alpha_t + \beta_t) - \Psi(\alpha_t) + \ln y_{i,t}) \times \alpha_t \\ (\Psi(\alpha_t + \beta_t) - \Psi(\beta_t) + \ln(1 - y_{i,t})) \times \beta_t \end{array} \right),
$$

and

$$
\mathcal{I}_t = K_t \times \left( \begin{array}{cc} \alpha_t^2(\Psi'(\alpha_t) + \Psi'(\alpha_t + \beta_t)) & -\alpha_t \beta_t \Psi'(\alpha_t + \beta_t) \\ -\alpha_t \beta_t \Psi'(\alpha_t + \beta_t) & \beta_t^2(\Psi'(\beta_t) + \Psi'(\alpha_t + \beta_t)) \end{array} \right),
$$

where $\Psi$ denotes the digamma function that is defined as $\Psi(x) = d \ln \Gamma(x)/d x$. We set the GAS scaling matrix to the inverse information matrix, $S_t = \mathcal{I}_t^{-1}$, to account for the curvature of
Table 1: Test statistics and critical values for the corporate LGD data, 1982Q1–2010Q1

<table>
<thead>
<tr>
<th>Stat</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td>LM GAS(0,1)</td>
<td>16.20</td>
<td>4.60</td>
<td>5.99</td>
</tr>
<tr>
<td>LM GAS(0,5)</td>
<td>22.84</td>
<td>15.99</td>
<td>18.31</td>
</tr>
<tr>
<td>MP10</td>
<td>-27.34</td>
<td>-12.80</td>
<td>-14.32</td>
</tr>
<tr>
<td>MP10*</td>
<td>-40.78</td>
<td>-12.80</td>
<td>-14.32</td>
</tr>
<tr>
<td>Andrews</td>
<td>18.56</td>
<td>10.01</td>
<td>11.79</td>
</tr>
<tr>
<td>Nyblom</td>
<td>1.63</td>
<td>0.61</td>
<td>0.75</td>
</tr>
</tbody>
</table>

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 full sample since the number of observations for each period is constant. In our setting of a varying number of observations $K_t$, we treat $K_t$ as given but multiply the Hessian in the algorithm of MP10 at time $t$ by $K_t/\bar{K}$, where $\bar{K}$ is the average of $K_t$ in the full sample. This modification follows from a similar derivation that is used for the information matrix in (17). It corrects the steps in the MP10 algorithm for periods when there are either many or few LGD observations in the cross section.

Next 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 the GAS(1,1) model as given by

$$f_{t+1} = \omega + A s_t + B f_t, \quad s_t = S_t \nabla_t, \quad A = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_{11} & 0 \\ 0 & b_{22} \end{pmatrix}. \tag{18}$$

The model with full rather than diagonal matrices for $A$ and $B$ produces similar results and is therefore omitted. The parameter estimates are presented in Table 2. All parameter estimates are strongly significant, except $\omega_2$. We find that there is strong persistence in both $\alpha_t$ and $\beta_t$, as both $b_{11}$ and $b_{22}$ are relatively high. Interestingly, the persistence in $\alpha_t$ is not as strong as that in $\beta_t$, that is $b_{11} < b_{22}$. Since $\alpha_t$ and $\beta_t$ characterize the mean of the beta distribution when it is close to 0 and 1, respectively, the higher persistence of $\beta_t$ indicates that the higher LGDs are more persistent than low LGDs. Such differences do not appear between $a_{11}$ and $a_{22}$.

Our second estimate of $f_t$ is obtained as a by-product of the MP10 algorithm. It is based on the Weighted Average Risk estimate of the path $f_t$ for several local alternatives as explained...
Table 2: GAS(1, 1) coefficients estimation results

<table>
<thead>
<tr>
<th>Coeff.</th>
<th>Estimate</th>
<th>Std Err</th>
<th>t–stat</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\omega_1$</td>
<td>0.0559 (0.0088)</td>
<td>6.3449</td>
<td></td>
</tr>
<tr>
<td>$\omega_2$</td>
<td>-0.0023 (0.0073)</td>
<td>-0.3075</td>
<td></td>
</tr>
<tr>
<td>$a_{11}$</td>
<td>0.1943 (0.0220)</td>
<td>8.8156</td>
<td></td>
</tr>
<tr>
<td>$a_{22}$</td>
<td>0.1836 (0.0361)</td>
<td>5.0770</td>
<td></td>
</tr>
<tr>
<td>$b_{11}$</td>
<td>0.8571 (0.0233)</td>
<td>36.8217</td>
<td></td>
</tr>
<tr>
<td>$b_{22}$</td>
<td>0.9235 (0.0355)</td>
<td>26.0252</td>
<td></td>
</tr>
</tbody>
</table>

in Section 2.2. 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 presented in Figure 5.

The LGD observations range from close to zero to almost one for given cross sections. In Figure 5 we also plotted the mean of the beta distribution $\alpha_t/(\alpha_t + \beta_t)$. 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. We point out that the GAS estimate is a filter (produces a one-sided estimate) and the MP estimate is a smoother (produces a two-sided estimate). In the latter case, future observations are also taken into account. We also observe that the two estimates differ substantially in the period before the 2008 financial crisis. The GAS estimate reveals a more moderate trough than the MP estimate.

We further conclude that the MP estimate is rather successful in extracting the mean signal throughout the sample while it is designed for local time variation only. The smoothed path of $f_t$ in Figure 5 for MP10* is constructed by a weighted average of 10 different paths, corresponding to the autoregressive coefficients $b_{11} = b_{22} = 1 - c/T$ for $c = 0, 5, 10, \ldots, 50$, with $T = 113$. The largest weights are assigned to the paths corresponding to $c = 30, \ldots, 50$ while the mode weight is at $c = 40$; it corresponds to an autoregressive decay of $1 - 40/113 \approx 0.65$. This autoregressive coefficient is lower than those of the GAS model, $b_{11}$ and $b_{22}$, in Table 2. Moreover, the autoregressive coefficient in the MP10 method is the same for $\alpha_t$ and $\beta_t$, in contrast to the GAS model. A smaller persistence parameter in MP10* is counterbalanced by a higher innovation variance in order to match the unconditional variance. The two effects lead to an MP10* estimate that is more sensitive to the small LGD values in the period leading up to the 2008 credit crisis. We emphasize that the MP10* test is not influenced by the less persistent paths. The MP10* test is based on the local alternative $c = 10$ that corresponds to...
Figure 5: Corporate Loss Given Default (LGD) data, 1982–2010
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).

The right hand panel of Figure 5 presents the estimates of the variance of our beta model as given by $\alpha_t / ((\alpha_t + \beta_t)^2 (1 + \alpha_t + \beta_t))$. The time-varying variance is slightly trending upwards. The variation in more recent LGD percentages is somewhat larger than in the early 1980s. We observe two peaks in the variance. These are linked to periods when LGD observations are sparse and the corresponding relative dispersions are high. The variance estimates in the MP10* and GAS frameworks are roughly similar. The main differences are in the periods around 1997 and around 2004–2006. In the latter period we find that the lower mean for MP10* in the left panel of Figure 5 is partly compensated by the higher variance. The MP10* test is not affected because it is based on the autoregressive coefficient of approximately 0.91 for MP10*, rather than 0.65 (implied by the mode $c = 40$) for the smoothed estimates reported in Figure 5.

5 Conclusions

We have proposed a new omnibus misspecification test for parameter instability in general nonlinear non-Gaussian time series models. By adopting the generalized autoregressive score, or GAS, model of Creal et al. (2013) as our specification of time-varying parameters, we have proposed a Lagrange Multiplier test for the null of constant parameters against the time-varying
alternative. We have carried out an extensive Monte Carlo study to investigate the finite sample
properties of the new test and have compared it with other competing general purpose tests.
Each of these tests are based on different time-varying parameter frameworks: the structural
breaks of Andrews (1993), the local parameter-driven variations of Müller and Petalas (2010),
and the martingale processes of Nyblom (1989).

We have concluded that the new test has robust power performance. For different time-
varying parameter processes, the power of the GAS test remains relatively constant, whereas
the power 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 consistent with the theory developed in their paper.

We have applied our tests to an empirical panel data set consisting of loss given default
percentages of corporate bonds. We have shown how the tests can be used in a practical
setting and how the tests can be adapted in cases with a time-varying number of observations.
Interestingly, we have 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 paths from their
algorithm produce similar results as the estimated path from the GAS model in our empirical
application. It illustrates that the two testing paradigms can provide complementary as well
as mutually reinforcing evidence in empirical studies.

References


Andrews, D. W. K. and W. Ploberger (1994). Optimal tests when a nuisance parameter is present only under


Bauwens, L. and D. Veredas (2004). The stochastic conditional duration model: A latent factor model for the


