Testing Structural Stability in Macroeconometric Models[1]

Otilia Boldea
Tilburg University

and

Alastair R. Hall[2]
University of Manchester

December 6, 2011

[1]This manuscript is prepared for Handbook of Research Methods and Applications on Empirical Macroeconomics edited by N. Hashimzade and M. Thornton which is to be published by Edward Elgar. The second author acknowledges the support of the ESRC grant RES-062-23-1351. We thank Denise Osborn for useful conversations in the preparation of this manuscript.

[2]Corresponding author. Economics, SoSS, University of Manchester, Manchester M13 9PL, UK. Email: alastair.hall@manchester.ac.uk

1 Introduction

Since the earliest days of macroeconometric analysis, researchers have been concerned about the appropriateness of the assumption that model parameters remain constant over long periods of time; for example see Tinbergen (1939).[1] This concern is also central to the so-called the Lucas (1976) critique which has played a central role in shaping macroeconometric analysis in the last thirty years. Lucas (1976) emphasizes the fact that the decision models of economic agents are hard to describe in terms of stable parametrizations, simply because changes in policy may change these decision models and their respective parametrization. These arguments underscore the importance of using structural stability tests as diagnostic checks for macroeconometric models.

A large body of empirical macroeconomic studies provides evidence for parameter instability in a variety of macroeconomic models. For example, considerable evidence exists that the New Keynesian Phillips curve has become flat and/or less persistent in recent years - see e.g. Alogoskoufis and Smith (1991), Cogley and Sargent (2001), Zhang, Osborn, and Kim (2008), Kang, Kim, and Morley (2009). Similarly, there is evidence that the interest rate reaction function is asymmetric over the business cycle - see e.g. Boivin and Giannoni (2006), Surico (2007), Benati and Surico (2008), Liu, Waggoner, and Zha (2009). Examples of parameter instability are not confined to monetary policy, but also extend to: growth models - Ben-David, Lumsdaine, and Papell (2003); output models - Perron (1997), Hansen (1992); exchange rate models - Rossi (2006); unemployment rate models - Weber (1995), Papell, Murray, and Ghiblawi (2000), Hansen (1997b), and many more. If such instabilities are ignored in the estimation procedure, they lead to incorrect policy recommendations and flawed macroeconomic forecasts.

Thus, it is essential - and it has become common practice - to test for instability in macroeconomic models. Instabilities can be of many types. In this chapter, we describe econometric tests for three main types of instability: *parameter breaks*, *other parameter instabilities* and *model instabilities*.

The first category, *parameter breaks*, focuses on sudden parameter changes. It may be desired to test for change that occurs at a particular time. In this case, the time of change, called *break-point* in econometrics and *change-point* in statistics, is said to be known. It can also be that it is desired to test for change at some unspecified point in the sample. In this case, the break-point is said to be unknown. We discuss tests for both a known and an unknown break. We also present methods for detecting multiple break-points, and their practical implementation.

For the second category, *other parameter instabilities*, we give a brief but thorough account of the state of the art tests for threshold models, smooth transition models and Markov-switching models.

In practice, there is no reason to assume that only parameters change, while the underlying functional form stays the same. Therefore, the third category

[1] See Morgan (1990) for review of Tinbergen's contributions.

of tests we describe is for *model instabilities*, when the functional form of the model is allowed to change after a known or unknown break.

In macroeconomics, the tests we present are often viewed as tests for Lucas critique. Lucas critique, at the time it was written, was directed against the use of backward-looking expectations, which did not take into account that agents will change their decision making when policy changes occur, leading to parameter or model instability. However, as Estrella and Fuhrer (2003) point out, using forward-looking expectations that incorporate certain policy changes does not make a model immune either to Lucas critique or to parameter or model instability. To see this, note that even if a structural model of the economy exists - with forward-looking behavior, based on solid microeconomic foundations, with important policy functions specified - it is unclear one can derive the complete model, or estimate it in practice. Any simplification, parameter calibration, omission or misspecification of relevant policy functions and other agent decisions can lead to parameter or model instability. Thus, even though it is often stated that Lucas critique implies that only reduced-form models suffer from instability - see e.g. Lubik and Surico (2010), in practice all models are prone to this problem and need to be tested for instability.[2]

In this chapter, we discuss Wald-type tests for breaks that are based on least-squares (LS) type methods, suitable for reduced-form models, and also on two-stage least-squares (2SLS) and generalized method of moments (GMM) estimation, more suitable for structural models. The chapter is organized as follows. In Section 2, we present structural stability tests for a single break based on GMM estimation. In Section 3, we discuss testing strategies for multiple break-points. Section 4 provides an brief but thorough account of tests for other types of parameter instability, with comments on the most recent developments in this literature. Section 5 focuses on testing for model instabilities rather than parameter instabilities. Section 6 concludes.

2 Testing for discrete parameter change at a single point

In this section, we summarize the literature on testing for discrete parameter change in macroeconometric models based on GMM. GMM provides a method for estimation of the parameters of a macroeconomic model based on the information in a population moment condition. GMM is described elsewhere in this volume, see Chapter 14, and so here we assume knowledge of the basic GMM framework. For ease of reference, we adopt the same generic notation as in Hall

[2]The name "structural instability" refers to the economic structure being unstable, encompassing instability in both structural and reduced-form models.

(2011): thus, the population moment condition is written as $E[f(v_t, \theta_0)] = 0$ in which θ_0 denotes the true value of the $p \times 1$ parameter, v_t is a random vector, and $f(\cdot)$ is a $q \times 1$ vector of continuous differentiable functions.[3] The sample is assumed to consist of observations on $\{v_t,\ t = 1, 2, \ldots T\}$.

As befits the GMM framework, the null and alternative hypotheses are expressed in terms of the population moment condition. The null hypothesis is that the population moment condition holds at the same parameter value throughout the sample. The alternative of interest here is that the population moment condition holds at one value of the parameters up until a particular point in the sample, called break-point, after which the population moment condition holds at a different parameter value. To formally present the hypotheses of interest here, we need a notation for the break-point. Following the convention in the literature, we let λ be a constant defined on $(0, 1)$ and let $[T\lambda]$ denote the potential break-point at which some aspect of the model changes.[4] For our purposes here, it is convenient to divide the original sample into two sub–samples. Sub–sample 1 consists of the observations before the break-point, namely $\mathcal{T}_1(\lambda) = \{1, 2, \ldots, [\lambda T]\}$, and sub–sample 2 consists of the observations after the break-point, $\mathcal{T}_2(\lambda) = \{[\lambda T] + 1, \ldots T\}$.

As mentioned in the introduction, this break-point may be treated as *known* or *unknown* in the construction of the tests. If it is known, then the break-point is specified *a priori* by the researcher and it is only desired to test for instability at this point alone. For example, Clarida, Gali, and Gertler (2000) investigate whether the monetary policy reaction function of the Federal Reserve Board is different during the tenure of different chairmen. Of particular interest in their study is whether or not the the reaction function is different pre- and post-1979, the year Paul Volcker was appointed as Chairman. Since their analysis uses quarterly data, this involves exploring whether or not there is parameter change at the fixed break date with $[T\lambda] = 1979.2$. If the break-point is unknown, the alternative is the broader hypothesis that there is parameter change at some point in the sample. We begin our discussion with the simpler case in which the break-point is known, and then consider the extension to the unknown break-point case.

For a fixed break-point indexed by λ, the null hypothesis of interest can be expressed mathematically as,

$$H_0:\ E[f(v_t, \theta_0)] \ = \ 0, \qquad \text{for } t = 1, 2, \ldots, T, \tag{1}$$

and the alternative hypothesis as,

$$H_1(\lambda): \quad \begin{array}{rcll} E[f(v_t, \theta_1)] & = & 0, & \text{for } t \in \mathcal{T}_1(\lambda) \\ E[f(v_t, \theta_2)] & = & 0, & \text{for } t \in \mathcal{T}_2(\lambda), \end{array}$$

where $\theta_1 \neq \theta_2$.

[3] We adopt the convention of using "0" to denote either a scalar, vector or matrix of zero(s) with the dimension determined by the term on the other side of the equation.

[4] Here, $[\cdot]$ stands for the integer part.

This statement of the alternative allows all elements of the parameter vector to change; this scenario is referred to as "pure structural change" in the literature. It is also possible to restrict attention to an alternative in which only certain elements of the parameter vector are allowed to change with the remainder taking the same value before and after the break-point; this scenario is referred to as "partial structural change". Given space limitations, we focus on the case of pure structural change which is arguably of most practical interest. We note parenthetically that all the methods discussed can be adapted to test for partial structural change in a relatively straightforward fashion, and that the qualitative discussion of the properties of the tests below also extends to those for partial structural change.

Andrews and Fair (1988) propose Wald, Lagrange Multiplier (LM) and Difference (D) statistics to test H_0 versus $H_1(\lambda)$. These statistics are actually derived by applying the test principles concerned to a different but equivalent specification of the null and alternative hypothesis in terms of a set of linear restrictions on an augmented parameter vector that indexes an augmented population moment condition. Since all three statistics have the same limiting properties under null and local alternatives, we focus exclusively on the Wald statistic and use it to discuss issues common to all three tests. As emerges below, the Wald statistic has an appealing intuitive structure that can be motivated without appealing to Andrews and Fair's (1988) framework involving augmented parameter vectors and moments; we therefore do not describe their approach, leaving the interested reader to refer to their paper or Hall (2005)[Chap 5.4].

The form of the Wald statistic can be motivated as follows. The null hypothesis is that the parameters take the same value before and after the break-point, and the alternative hypothesis is that the parameters take a different value before and after the break-point. Given this structure, a natural way to assess which is true is to estimate the parameters based on the observations in $\mathcal{T}_1(\lambda)$ and $\mathcal{T}_2(\lambda)$ separately and then compare these estimators. If the null is true there should be no difference between them - allowing for sampling variation - while if the alternative is true, there should be a difference. This is the essence of the Wald statistic.

To present the formula for the statistic, we require certain notation. For $i = 1, 2$, let $g_{i,T}(\theta; \lambda) = T^{-1} \sum_{t \in \mathcal{T}_i(\lambda)} f(v_t, \theta)$, where $\sum_{t \in \mathcal{T}_i(\lambda)}$ denotes summation over t for all values in $\mathcal{T}_i(\lambda)$, $G_{i,T}(\theta; \lambda) = \partial g_{i,T}(\theta; \lambda)/\partial \theta'$, $S_i(\lambda) = lim_{T \to \infty} Var[T^{1/2} g_{i,T}(\theta_i)]$ and $S_{i,T}(\lambda)$ be a consistent estimator for $S_i(\lambda)$. The sub-sample parameter estimators referred to in the previous paragraph are calculated in the following way.

Definition 1 *For $i = 1, 2$, $\hat{\theta}_{i,T}(\lambda)$ is defined to be the GMM estimator of θ_i based on the population moment condition $E[f(v_t, \theta_i)] = 0$ calculated from observations in $\mathcal{T}_i(\lambda)$ and using weighting matrix $W_{i,T} = \{S_{i,T}(\lambda)\}^{-1}$.*

In the literature, $\hat{\theta}_{i,T}(\lambda)$, $i = 1, 2$, are often referred to as "partial-sum" GMM estimators because they are based on the part of the sample, either up to or after the break-point. Notice that the specified choice of weighting matrix is optimal, and that in practice the estimators would be obtained using a two-step or iterated estimation; see Chapter 14 for further details.

The Wald test statistic is as follows:

$$\mathcal{W}_T(\lambda) \;=\; T\left[\hat{\theta}_{1,T}(\lambda) \;-\; \hat{\theta}_{2,T}(\lambda)\right]' \hat{V}_{\mathcal{W}}(\lambda)^{-1} \left[\hat{\theta}_{1,T}(\lambda) \;-\; \hat{\theta}_{2,T}(\lambda)\right], \qquad (2)$$

where, using $\hat{\theta}_i = \hat{\theta}_{i,T}(\lambda)$ for ease of notation,

$$\begin{aligned} \hat{V}_{\mathcal{W}}(\lambda) \;&=\; \frac{1}{\lambda}[G_{1,T}(\hat{\theta}_1;\lambda)'W_{1,T}(\lambda)G_{1,T}(\hat{\theta}_1;\lambda)]^{-1} \;+ \\ &\quad \frac{1}{(1-\lambda)}[G_{2,T}(\hat{\theta}_2;\lambda)'W_{2,T}(\lambda)G_{2,T}(\hat{\theta}_2;\lambda)]^{-1}. \end{aligned} \qquad (3)$$

The limiting distribution of the Wald statistic under the null hypothesis is given in the following proposition.

Proposition 1 *If certain regularity conditions hold then under* H_0, $\mathcal{W}_T(\lambda) \xrightarrow{d} \chi^2_p$.

The regularity conditions referred to in Proposition 1 are the same as those needed for the standard first order asymptotic theory of GMM estimators; see Andrews and Fair (1988) or Hall (2005)[Chap. 5.4]. They include the crucial assumptions that v_t is non-trending, the function $f(v_t, \cdot)$ is smooth in θ, and that the parameters are identified.[5]

As anticipated in our motivation of the test above, the test statistic behaves very differently under $H_1(\lambda)$. For in this case, we have

$$\hat{\theta}_{1,T}(\lambda) \;-\; \hat{\theta}_{2,T}(\lambda) \;\xrightarrow{p}\; \theta_1 - \theta_2 \;=\; \mu(\lambda), \text{ say,}$$

with $\mu(\lambda) \neq 0$ because $\theta_1 \neq \theta_2$. As a result, it can be shown that $\mathcal{W}_T(\lambda)$ diverges as $T \to \infty$; in consequence of which $\mathcal{W}_T(\lambda)$ is said to be a *consistent* test of $H_0(\lambda)$ versus $H_1(\lambda)$.[6]

While the statistic is designed to test for parameter change at a particular point in the sample and is consistent against that alternative, caution needs to be exercised in interpreting the outcome of the test. To see why, we now consider the behaviour of $\mathcal{W}_T(\lambda)$ if the parameters do not change at $T\lambda$ but

[5] In instrumental variables estimation, identification is commonly referred to as the case of "strong instruments". For break-point tests in the presence of weak instruments, see Caner (2011).

[6] See Andrews and Fair (1988) for a local power analysis of the test.

do change at some other point in the sample. Let this true break-point be indexed by λ^* and, for sake of argument, assume $\lambda^* > \lambda$ that is, $\mathcal{W}_T(\lambda)$ is designed to test for parameter change at a point in the sample before it actually occurs. Further assume the population moment condition is satisfied at θ_1^* for observations $t \leq [T\lambda^*]$, and is satisfied at θ_2^* for observations $t > [T\lambda^*]$. In this case, $\mathcal{T}_1(\lambda)$ contains observations for which $E[f(v_t, \theta_1^*)] = 0$, and so $\hat{\theta}_{1,T}(\lambda) \xrightarrow{p} \theta_1^*$; but $\mathcal{T}_2(\lambda)$ contains some observations for which $E[f(v_t, \theta_1^*)] = 0$ and some for which $E[f(v_t, \theta_2^*)] = 0$, and, as a result, it can be shown that $\hat{\theta}_{2,T}(\lambda) \xrightarrow{p} h(\theta_1^*, \theta_2^*)$ for some function $h(\,\cdot\,,\,\cdot\,)$. Therefore, under this scenario, we have

$$\hat{\theta}_{1,T}(\lambda) - \hat{\theta}_{2,T}(\lambda) \xrightarrow{p} \theta_1^* - h(\theta_1^*, \theta_2^*) = \mu^*(\lambda, \lambda^*), \text{ say.}$$

In general, $\mu^*(\lambda, \lambda^*) \neq 0$, and so by similar arguments to above the test diverges in this case as well.[7] Thus in the limit, the test rejects with probability one when T is large even if the wrong break-point has been specified under the alternative. In finite samples, the impact is less clear because - loosely speaking - $h(\theta_1^*, \theta_2^*)$ is a weighted average of θ_1^* and θ_2^*: if $h(\theta_1^*, \theta_2^*)$ is close to θ_2^* then the test is more likely to reject than if $h(\theta_1^*, \theta_2^*)$ is close to θ_1^* *ceteris paribus*. Either way, the possibility of parameter change at other points in the sample complicates the interpretation of the outcome of the fixed break-point test, and motivates the use of the unknown break-point tests to which we now turn.

If the break-point is unknown, then it is desired to test whether there is evidence of instability at any point in the sample. However, in practice, it is necessary to limit attention to potential breaks indexed by λ values within a closed subset of the unit interval that is, $\lambda \in \Lambda = [\lambda_l, \lambda_u] \subset [0, 1]$. The choice of Λ is critical and typically governed by two main considerations: on the one hand, given the alternative of interest, it is desirable for Λ to be as wide as possible; on the other hand, it must not be so wide that asymptotic theory is a poor approximation in the sub–samples. In applications to models of economic time series, it has become customary to use Λ equal to $[0.15, 0.85]$ or (less often) $[0.20, 0.80]$. The null hypothesis is again H_0 in (1). The alternative is

$$H_1(\Lambda) = H_1(\lambda) \text{ holds for some } \lambda \in \Lambda.$$

The construction of statistics for testing H_0 versus $H_1(\Lambda)$ is a natural extension of the fixed break-point methods. In this setting, $\mathcal{W}_T(\lambda)$ is calculated for each possible λ to produce a sequence of statistics indexed by λ, and inference is based on some function of this sequence.[8] Three functions of this sequence have become popular in the literature and these lead to the so-called "Sup-",

[7] See Hall and Sen (1999)[Theorem 3.2] and Sen (1997) for a similar analysis using local alternatives.

[8] Inference can also be based on the D or LM statistics mentioned above and the discussion below equally applies to these statistics as well.

"Av-" and "Exp-" statistics which are respectively given by,[9]

$$\begin{aligned} Sup\mathcal{W}_T &= \sup_{\lambda\in\Lambda}\{\,\mathcal{W}_T(\lambda)\,\}, \\ Av\mathcal{W}_T &= \int_{\Lambda}\mathcal{W}_T(\lambda)dJ(\lambda), \\ Exp\mathcal{W}_T &= log\left\{\int_{\Lambda} exp[0.5\mathcal{W}_T(\lambda)]dJ(\lambda)\right\}, \end{aligned}$$

where $J(\lambda) = (\lambda_u - \lambda_l)^{-1}d\lambda$. As they stand these statistics are not operational because we have treated λ as continuous, whereas in practice it is discrete. For a given sample size, the set of possible break-points are $\mathcal{T}_b = \{i/T;\ i = [\lambda_l T], [\lambda_l T]+1, \ldots, [\lambda_u T]\}$. So in practice, inference is based on the discrete analogs to $Sup\mathcal{W}_T$, $Av\mathcal{W}_T$ and $Exp\mathcal{W}_T$:

$$\begin{aligned} Sup\mathcal{W}_T &= \sup_{i\in\mathcal{T}_b}\{\,\mathcal{W}_T(i/T)\,\} \\ Av\mathcal{W}_T &= d_b^{-1}\sum_{i=[\lambda_l T]}^{[\lambda_u T]}\mathcal{W}_T(i/T) \\ Exp\mathcal{W}_T &= log\left\{d_b^{-1}\sum_{i=[\lambda_l T]}^{[\lambda_u T]} exp[0.5\mathcal{W}_T(i/T)]\right\} \end{aligned}$$

where $d_b = [\lambda_u T] - [\lambda_l T] + 1$. Various statistical arguments can be made to justify one statistic over another, but it has become common practice to report all three in the empirical literature.

The limiting distribution of the three statistics is given in the following proposition.[10]

Proposition 2 *If certain regularity conditions hold then under H_0, then we have: $Sup\mathcal{W}_T \Rightarrow Sup_{\lambda\in\Lambda}\mathcal{W}(\lambda)$, $Av\mathcal{W}_T \Rightarrow \int_{\Lambda}\mathcal{W}(\lambda)dJ(\lambda)$, and $Exp\mathcal{W}_T \Rightarrow$*

[9] This function is chosen to maximize power against a local alternative in which a weighting distribution is used to indicate the relative importance of departures from parameter constancy in different directions (*i.e.* $\theta_1 - \theta_2$) at different break-points and also the relative importance of different break-points; see Andrews and Ploberger (1994) and Sowell (1996). The distribution over break-points is commonly taken to be uniform on Λ which is imposed in the presented formulae (via the specified $J(\pi)$) for the Av- and Exp- statistics in the text.

[10] For regularity conditions and proofs see: Sup-test, Andrews (1993); Av-, Exp- tests, Andrews and Ploberger (1994) (in context of maximum likelihood) and Sowell (1996) (in context of GMM). In this context "⇒" denotes weak convergence in distribution.

$log \left[\int_{\Lambda} exp\{0.5\mathcal{W}(\lambda)\} dJ(\lambda) \right]$ *where* $\mathcal{W}(\lambda) = \{\lambda(1-\lambda)\}^{-1} BB_p(\lambda)' BB_p(\lambda)$ *and* $BB_p(\lambda)$ *denotes a* $p \times 1$ *Brownian Bridge on* $[0,1]$.

The limiting distributions in Proposition 2 are non-standard but depend only on p, the dimension of the parameter vector. Percentiles are reported in Andrews (2003) (Sup-) and Andrews and Ploberger (1994) (Av- and Exp-). These percentiles enable the researcher to ascertain whether the statistic is significant at a prescribed level. Hansen (1997a) reports response surfaces which can be used to calculate approximate p–values for all three versions of these tests.

The use of unknown break-point tests removes the concern about the interpretation of fixed break-point tests, but the tests are still only designed against an alternative in which there is one break-point. This is a limitation because in many cases of interest there are likely to be multiple events in the sample which may have caused the parameters to change. In the next section, we describe various tests that can be used to test for parameter change at multiple unknown break-points.

3 Testing discrete parameter change at multiple points

Tests for multiple break-points have been proposed predominantly in the context of models that can be estimated via LS-type criteria, such as ordinary least-squares (OLS), two-stage least-squares (2SLS) and nonlinear least-squares (NLS). The most widely used testing strategy for multiple breaks in linear models estimated via ordinary least-squares (OLS) is the one proposed by Bai and Perron (1998). This strategy involves three types of tests: (i) testing no breaks versus a known number of breaks; (ii) testing no breaks against an unknown number of breaks up to a fixed upper bound, and (iii) testing ℓ versus $\ell+1$ breaks.

These tests are useful as their by-products are consistent estimates for the break locations - see Bai and Perron (1998). The strategy for determining the number of break-points in a sample involves, as a first step, testing zero versus a known or unknown number of breaks, via tests in (i)-(ii), described below. It is common to test for maximum five breaks. If the null of zero breaks is not rejected, we conclude that there are no breaks. If the null is rejected, it implies we have at least one break, and so we employ the tests in (iii), described below, for one versus two breaks. If we do not reject, then we conclude we have one break; if we reject, then we have at least two breaks and test via the tests in (iii) for two versus three breaks. If we do not reject, then we conclude we have two breaks. If we reject, then we continue testing for an additional break until we cannot reject the null or a maximum number of breaks has been reached.

This is a simple sequential strategy for estimating the number of breaks, and provided the significance level of each test is shrunk in each step towards zero[11], we will obtain the true number of breaks with probability one for large T.

For describing these tests, consider the following univariate linear model, estimable via OLS:

$$y_t \;=\; x_t'\theta_i + u_t \qquad (t = T_{i-1}^0 + 1, ..., T_i^0) \qquad (i = 1, \dots, m+1) \tag{4}$$

where y_t is a scalar dependent variable, x_t is a $p \times 1$ vector of exogenous regressors, uncorrelated with u_t, possibly including lags of y_t. Also, the number of breaks m is fixed, $T_i^0 = [\lambda_i^0 T]$ are the true break-points, λ_i^0 are the true break-fractions, for $i = 0, \dots, m+1$, and $\lambda_0^0 = 0, \lambda_{m+1}^0 = 1$ by convention. Here, we treat λ_i^0 as unknown quantities; the tests for unknown break-points simplify as in the previous section when the λ_i^0's are known.

(i) Tests for a fixed number of breaks

Under the notation above, the tests for a fixed number of breaks are for the following null and alternative hypotheses:

$$H_0 : \; m = 0 \qquad H_1 : \; m = k, \text{for a fixed } k. \tag{5}$$

Since the LM and LR tests are asymptotically equivalent to the Wald test, we restrict our attention to the latter. To derive Wald tests from their principles, rewrite the null hypothesis in terms of restricting the parameters to be the same across sub-samples:

$$H_0 : \theta_1 = \theta_2 = \ldots = \theta_{k+1} \tag{6}$$

$$H_1 : \theta_i \neq \theta_j \text{ for all } i \neq j, \qquad (i, j = 1, \dots, k+1). \tag{7}$$

To define the Wald test, let $\tilde{R}_k$ be the $k \times (k+1)$ matrix with the $(j,j)^{th}$ element equal to 1, the $(j, j+1)^{th}$ element equal to -1, and the rest equal to zero, for $j = 1, 2, \dots, k$. Also let $R = \tilde{R} \otimes I_p$, where $\otimes$ denotes the Kronecker product. Then H_0 above can be written as $R_k\theta^c = 0$, where θ^c is a $(k+1)p$ vector that vertically stacks $\theta_1, \dots, \theta_{k+1}$. Let $\lambda^c = (0, \lambda_1, \lambda_2, \dots, \lambda_k, 1)'$ the break-fractions associated with any candidate partition of the sample $T^c = (0, T_1, T_2, \dots, T_k, T)'$, where $T_i = [\lambda_i T]$ for $i = 1, 2, \dots, k$. Also, let the corresponding sub-samples be denoted by $\mathcal{T}_i(\lambda^c) = \{[\lambda_{i-1}T]+1, [\lambda_{i-1}T]+2, \dots, [\lambda_i T]+1\}$ for $i = 1, 2, \dots, k+1$. Since the Wald tests are constructed using estimates under the alternative, we have to estimate $\theta_1, \dots, \theta_{k+1}$.

Definition 2 *For $i = 1, 2, \dots, k+1$, $\hat{\theta}_{i,T}(\lambda^c)$ is defined to be the OLS estimator of θ_i, based on minimizing the sum of squared residuals calculated from*

[11] For details, see Bai and Perron (1998).

observations in $\mathcal{T}_i(\lambda^c)$*. Also,* $\hat{\theta}^c_{i,T}(\lambda^c)$ *is the* $(k+1)p$ *vector that vertically stacks* $\hat{\theta}_{i,T}(\lambda^c)$*, for* $i = 1, \ldots, m+1$*.*

Let also $V_i(\lambda^c) = lim_{T\to\infty} Var(T^{1/2}\,[\hat{\theta}_{i,T}(\lambda) - \theta_i])$, $V_{i,T}(\lambda^c)$ be a consistent estimator of $V_i(\lambda^c)$ and $V_T(\lambda^c)$ the $(k+1)\times(k+1)$ block-diagonal matrix with the $(i,i)^{th}$ diagonal block equal to $V_{i,T}(\lambda^c)$. Then the Wald test for a particular sample partition λ^c is defined as follows:

$$\mathcal{W}_T(\lambda^c) \;=\; T[R_k\,\hat{\theta}^c_{i,T}(\lambda^c)]'\,[R_k\,V_T(\lambda^c)\,R_k']^{-1}[R_k\,\hat{\theta}^c_{i,T}(\lambda^c)] \tag{8}$$

For $k=1$, this test is the OLS equivalent of its GMM counterpart in equation (2). As in the previous section, it depends on the particular partition of the sample used, so for unknown break-points, we use its $sup\mathcal{W}_T$ version, defined as:

$$Sup\mathcal{W}_T(k) = \sup_{\lambda^c\in\Lambda^c_\epsilon}\{\,\mathcal{W}_T(\lambda^c)\,\}, \tag{9}$$

where $\Lambda^c_\epsilon = \{(\lambda_1,\lambda_2,\ldots,\lambda_{k+1}) : |\lambda_i - \lambda_{i-1}| > \epsilon,\ (i=1,\ldots,k+1)\}$ for some positive ϵ. In practice, ϵ is usually chosen to be 0.15; this implies not only that the breaks cannot be too close to the beginning or end of the sample, but also that there are enough observations in each sub-sample so that OLS estimation can be performed. The asymptotic distribution of this test is given below.

Proposition 3 *Under certain regularity conditions and* H_0 *in* (6)*,*

$Sup\mathcal{W}_T(k) \Rightarrow \sup_{\lambda^c\in\Lambda^c_\epsilon}\mathcal{W}(\lambda^c)$*, where*

$$\mathcal{W}(\lambda^c) = B'_{p(k+1)}\,\{\,[C_k^{-1}\tilde{R}_k'(\tilde{R}_kC_k^{-1}\tilde{R}_k')^{-1}\tilde{R}_kC_k^{-1}]\otimes I_p\,\}\,B_{p(k+1)},$$

with $B_{p(k+1)} = [B_p'(\lambda_1), B_p'(\lambda_2) - B_p'(\lambda_1), \ldots, B_p'(\lambda_{k+1}) - B_p'(\lambda_k)]'$*, a* $p(k+1)\times 1$ *vector of pairwise independent vector of Brownian motion increments of dimensions* p*,* C_k *is a* $k\times k$ *diagonal matrix with elements* $\lambda_1, \lambda_2-\lambda_1, \ldots, \lambda_{k+1}-\lambda_k$ *on the diagonal, and* $\lambda_{k+1} = 1$ *by convention.*

The most important regularity conditions in Proposition 3 are that the regressors are not trending, they are orthogonal to the errors, that there is no unit root, and that there are no changes in the marginal distribution of x_t.[12] Unlike

[12] For tests that allow for changes in the marginal distribution of regressors, see Hansen (2000).

for LR-type tests, these regularity conditions allow for heteroskedasticity and autocorrelation. In particular, they allow for the variance of u_t to change at the same time as the parameters. As pointed out by Lubik and Surico (2010), this is an important feature of the $Sup\mathcal{W}_T(k)$ test that allows practitioners to test more accurately for monetary policy breaks during the Great Moderation.

The $Sup\mathcal{W}_T(k)$ test is consistent against H_1. For $k = 1$, its distribution reduces to the one in Proposition 1, and its optimality properties are the same as for GMM settings[13], but they are not known for $k > 1$. However, for most practical purposes, it suffices to know that in the OLS setting, the $\sup \mathcal{W}_T(k)$ delivers consistent estimators of the true k break-fractions indexing the break-points.

(ii) Tests for an unknown number of breaks

The test in (9) also rejects with probability one when T is large, if the true number of breaks under the alternative is $k^* \neq k$. However, if the true alternative is $H_1 : m = k^*$, in small samples it might not have good power properties because it is not designed for this alternative. To address this issue, Bai and Perron (1998) propose a second set of tests, presented in (ii) below, against the alternative of an unknown number of breaks up to a maximum:

$$H_0 : \ m = 0 \qquad H_1 \ : 1 \leq m \leq M, \text{for a fixed } M. \tag{10}$$

These tests are known as *double-maximum* or *Dmax*-type tests. The idea behind these tests is to construct for each $m \in \{1, \ldots, M\}$ a $sup\mathcal{W}_T(m)$-test (thus a maximum for each m), and then to maximize over weighted versions of these statistics to obtain a unique test statistic for the null and alternative hypotheses in (5). This test is defined below:

$$Dmax\mathcal{W}_T \ = \ \max_{1 \leq m \leq M} \frac{a_m}{p} \, Sup\mathcal{W}_T(m), \tag{11}$$

for some fixed, strictly positive weights a_m. The distribution of this test generalizes in a straightforward fashion from Proposition 3:

Proposition 4 *Under certain regularity conditions and H_0,*

$$Dmax\mathcal{W}_T \Rightarrow \max_{1 \leq m \leq M} \frac{a_m}{p} \sup_{\lambda^c \in \Lambda^c_\epsilon} \sum_{i=1}^{m+1} \mathcal{W}_i(\lambda^c)$$

[13] Corresponding $Ave\mathcal{W}_T$ and $Exp\mathcal{W}_T$ and their optimality properties for one break in an OLS setting are discussed in Kim and Perron (2009).

The regularity conditions are the same as for $Sup\mathcal{W}_T(m)$ test, for each m. The weights for the $Dmax\mathcal{W}_T$ test should be set larger for a certain m if one believes that m is more likely to be the correct number of breaks. If there is no clear a-priori belief about the true number of breaks, one can use equal weights $a_m = 1/M$, in which case the test is known as a $UDmax$ test. However, note that the scaling p is used because in its absence and with equal weights a_m, this test will be equivalent to testing zero against M breaks, since the critical values increase in m for a fixed p. Since, despite the scaling, the critical values still tend to increase with m, let $c(m, p, \alpha)$ be the asymptotic critical value of the test $Sup\mathcal{W}_T(m)/p$ at significance level $100\alpha\%$.[14] Then the problem of increasing critical values is alleviated by setting $a_1 = 1$ and $a_m = c(1, p, \alpha)/\, c(m, p, \alpha)$, and the corresponding test is called a $WDmax$ test.

(iii) Tests for an additional break

The third category of tests for multiple breaks (iii) are called *sequential Wald tests*, since they are Wald tests for an additional break. The null and alternative hypotheses are, for any given ℓ:

$$H_0 :\ m = \ell \qquad H_1\ : m = \ell + 1. \tag{12}$$

A LR-type test for is proposed in Bai and Perron (1998). However, asymptotically equivalent Wald-type tests can be derived as a special case of the sequential Wald tests in Boldea and Hall (2010) and Hall, Han, and Boldea (2012). For the Wald test, one ideally uses the estimates only under the alternative hypothesis.[15] However, for computational ease, it has become routine among practitioners to use Bai and Perron's (1998) approach of pre-estimating the model with ℓ breaks. This implies that estimates of the ℓ breaks are obtained as a by-product of calculating the Wald statistic $Sup\mathcal{W}_T(\ell)$, imposed as if they were the true ones, and for the alternative hypothesis in (3), evidence is maximized for exactly one additional break, occurring in only one of the $\ell + 1$ sub-samples obtained by partitioning the sample with the pre-estimated ℓ breaks. To define the test, let μ denote each candidate additional break fraction in the pre-estimated $\ell + 1$ sub-samples $\Lambda_q(q = 1, 2, \ldots, \ell + 1)$, with

$$\Lambda_q = \{\mu : \hat{T}_{q-1} + (\hat{T}_q - \hat{T}_{q-1})\eta \le [\mu T] \le \hat{T}_q - (\hat{T}_q - \hat{T}_{q-1})\eta\}.$$

In these sub-samples, the OLS parameter estimates before and after each candidate additional break are denoted by $\vartheta_q(\mu) = [\hat{\theta}_1(\mu, q)', \hat{\theta}_2(\mu, q)']'$, where $q = 1, 2, \ldots, \ell + 1$, and the restriction that they are equal is defined through the restriction matrix $R^* = [I_p; -I_p]$. Then the sequential Wald test is:

[14] These critical values can be found in Bai and Perron (1998).

[15] For an equivalent LR test that estimates the model in 4 with ℓ, respectively $\ell + 1$ breaks, see Bai (2009).

$$\mathcal{W}_T(\ell+1|\ell) \;=\; \max_{1\le q\le \ell+1}\, \sup_{\mu\in\Lambda_q} \mathcal{W}_{T,\ell}(\mu,q) \tag{13}$$

where

$$\mathcal{W}_{T,\ell}(\mu,q) \;=\; T\,[R^*\hat{\vartheta}_q(\mu)]'\,[R^* V_T^*(\mu,q) R^{*'}]^{-1}\, R^*\hat{\vartheta}_q(\mu)$$

and $V_T^*(\mu,q)$ is the $2p\times 2p$ block-diagonal matrix with diagonal blocks $V_{1,T}(\mu,q)$ and $V_{2,T}(\mu,q)$. The latter two are defined as consistent estimators of their asymptotic equivalents $V_1(\mu,q) = lim_{T\to\infty} Var\; T^{1/2}\,[\hat{\theta}_1(\mu,q)-\theta_q]$, respectively $V_2(\mu,q) = lim_{T\to\infty} Var\; T^{1/2}\,[\hat{\theta}_2(\mu,q)-\theta_q]$, for $q = 1,2,\ldots,\ell+1$.

The asymptotic distribution of the test is described below:

Proposition 5 *Under certain regularity conditions and H_0 in* (3)*,* $\lim P(\mathcal{W}_T(\ell+1|\ell)\le x) = G_{p,\eta}^{\ell+1}$*, where $G_{p,\eta}$ is the cumulative distribution function of*

$$\sup_{\eta\le\mu\le 1-\eta} \frac{\|B_p(\mu)-\mu B_p(1)\|^2}{\mu(1-\mu)}.$$

The regularity conditions follow directly from Hall, Han, and Boldea (2012), by treating all endogenous variables as exogenous. They allow for breaks in the error variance occurring at the same time as the parameters under H_1 in (3). Critical values for the tests (i)-(iii) can be found in Bai and Perron (1998), and p-values based on approximate response surfaces can be found in Hall and Sakkas (2011).

As mentioned above, these tests are useful for models that can be estimated by OLS, thus with exogenous regresssors. When some regressors are endogenous, Hall, Han, and Boldea (2012) show that a similar sequential procedure for finding the number of breaks in models with endogenous regressors can be developed, based on tests constructed with 2SLS estimates.

However, unlike for OLS, with 2SLS, one needs to first assess whether there are any breaks in the first-stage regression. To see why this is important, assume that the researcher has in mind an economic model, from which the first and second stages of 2SLS estimation arise naturally. For example, consider the following structural model:

$$y_t = \theta x_t + u_t \tag{14}$$

$$x_t = \gamma_1 y_t \;+\; \gamma_2 h_t \;+\; \gamma_3 z_{1,t} + v_{1,t} \tag{15}$$

$$h_t = \delta_1 x_t \;+\; \delta_2 z_{2,t} + v_{2,t} \tag{16}$$

where y_t, x_t, h_t are scalar dependent variables, $z_t = (z_{1,t}, z_{2,t})'$ are scalar exogenous regressors, $u_t, v_{1,t}, v_{1,t}$ are errors and $\theta, \gamma_1, \gamma_2, \gamma_3, \delta_1, \delta_2$ are scalar unknown parameters that may break at unknown locations in the sample.

If one is interested in estimating θ, the equation of interest is (14), and will be the second stage in a 2SLS regression, with the first stage instrumenting for the endogeneity of x_t via instruments z_t. In this example, the first-stage arises naturally, since the reduced form for x_t can be found by substituting (14) and (16) into (15):

$$
\begin{aligned}
x_t &= z_t'\Delta + v_t && (17)\\
\Delta &= \left(\frac{\gamma_3}{1-\theta\gamma_1-\delta_1\gamma_2}, \frac{\gamma_2\delta_2}{1-\theta\gamma_1-\delta_1\gamma_2}\right)' \\
v_t &= \frac{\gamma_1 u_t + v_{1,t} + \gamma_2 v_{2,t}}{1-\theta\gamma_1-\delta_1\gamma_2}.
\end{aligned}
$$

In this context, we can see that all breaks in θ, thus in (14), will also be in (17) by default, unless $\gamma_1 = 0$. When $\gamma_1 = 0$, if no other parameters change besides θ, (17) will have no breaks. If any of the parameters $\gamma_j (j = 1, 2, 3), \delta_1, \delta_2$ change, then these changes are only reflected in (17). Thus the first-stage can have no breaks, or breaks that are common to the second-stage, or breaks that are idiosyncratic to the first-stage.

In practice, one does not necessarily know which scenario occurs, so it is important to consider both the case where the first-stage is stable and where it is unstable. For simplicity, consider a data generating process with m and m^* breaks in the second-, respectively first-stage regressions:

$$y_t = x_t\theta_i + u_t \qquad (t = T^0_{i-1}+1, ..., T^0_i) \qquad (i = 1, \ldots, m+1) \tag{18}$$

where x_t is a scalar endogenous regressor, i.e. correlated with u_t, and thus needs to be predicted via the first stage OLS regression with $s \times 1$ strong instruments z_t:

$$x_t = z_t'\Delta_i + v_t \qquad (t = T^*_{j-1}+1, ..., T^*_j) \qquad (j = 1, \ldots, m^*+1) \tag{19}$$

with u_t correlated with v_t, $T_0 = T^*_0 = 1$, $T^0_i = [\lambda^0_i T]$, $T^*_j = [\lambda^*_j T]$, $T_{m+1} = T_{m^*+1} = T$, and some breaks may be common to both equations.

If there are no breaks in the first-stage, i.e. $m^* = 0$, the 2SLS structural stability tests are computed exactly as their OLS conterparts in (i)-(iii), but for the second-stage equation (18), and with x_t replaced by $\hat{x}_t$, its predicted counterpart from an OLS regression in (17). For clarity, the 2SLS estimators are defined below.

Definition 3 *For $i = 1, 2, \ldots, k+1$, $\hat{\theta}_{i,T}(\lambda^c)$ is defined to be the 2SLS estimator of θ_i, based on minimizing the OLS sum of squared residuals calculated for the second-stage* (18)*, from observations in $\mathcal{T}_i(\lambda^c)$ - defined as before- using as* *regressors $\hat{x}_t$ instead of x_t, where $\hat{x}_t$ is the full-sample OLS estimator from the first stage equation* (17). *Also, $\hat{\theta}^c_T(\lambda^c)$ is the $(k+1)p$ vector that vertically stacks $\hat{\theta}_{i,T}(\lambda^c)$, for $i = 1, \ldots, m+1$.*

(iv) Sequential testing strategy for stable first-stage regression

For sequential testing, the tests in (i)-(iii), $Sup\mathcal{W}_T$, $Dmax\mathcal{W}_T$ and $\mathcal{W}_T(\ell+1|\ell)$, are defined in the same way as before, except that the 2SLS estimators replace their OLS counterparts in the definition of the tests. As Hall, Han, and Boldea (2012) show, the asymptotic distributions of these tests are also the same as in Propositions 3-5. Thus, a 2SLS sequential strategy for estimating the number and location of breaks can be constructed in the same way from these tests as before, when the first-stage equation (19) is stable.

However, in general (19) may also have breaks, i.e. $m^* \neq 0$. One can test this equation for breaks, and find their locations, via the OLS sequential testing strategy in (i)-(iii). As discussed above, these tests also provide consistent estimators of the number of break-points, m^* and their locations, $T^*_j, (j = 1, 2, \ldots, m^*)$.

It remains to find the number of breaks in the second-stage, for which $\hat{x}_t$, a prediction of x_t based on estimating the first-stage equation (19), needs to be computed. If one ignores the breaks found in (19) in computing $\hat{x}_t$, the 2SLS tests in the second stage will pick up these breaks and reject with probability one for large T even if there are no breaks in the second-stage. If one computes $\hat{x}_t$ by OLS in each sub-sample constructed via the estimates of T^*_j, and then proceeds with testing for breaks in the full-sample of the second stage (18) via the tests in (i)-(iii), but with x_t replaced by $\hat{x}_t$, then the asymptotic distributions in Propositions 3-5 are no longer valid.[16]

Fortunately, there is a simple way to side-step these issues and sequentially test for breaks in (18). A strategy for finding these breaks is described below.

(v) Sequential testing strategy for unstable first-stage regression

(v-i) Tests for breaks that are idiosyncratic to the second-stage

If breaks are found in the first stage, a strategy for finding the breaks that only occur in the second stage is described below.

- Obtain estimates $\hat{m}^*$ for the number of breaks m^* and $\hat{T}^*_j (j = 1, \ldots, \hat{m}^*)$ for the associated break-points $T^*_j (j = 1, \ldots, m^*)$, either as a by-product of the OLS sequential strategy in (19), or by global estimation via the

[16] These distributions are much more complicated and depend on the relative positioning of the breaks in the first- and second-stage. To obtain critical values, bootstrap-based methods are proposed - see Boldea, Cornea, and Hall (2011).

methods in Hall, Han, and Boldea (2012). If the sequential strategy is used, reduce in each step the critical value to make sure that $\hat{m}^* = m^*$ with probability one as T grows larger. This ensures that we can treat $\hat{m}^*$ as if it were m^* in the next steps.

- Split the sample into sub-samples $\hat{\mathcal{T}}_j^* = \{\hat{T}_{j-1}^* + 1, \hat{T}_{j-1}^* + 2, \ldots, \hat{T}_j^*\}$ for $(j = 1, \ldots, \hat{m}^*)$. In each sub-sample $\hat{\mathcal{T}}_j^*$, compute $\hat{x}_t$ via OLS, and run the sequential testing strategy in (iv) for each of these sub-samples of the second-stage model (18).

- As a by-product of the testing strategy, or the re-estimation of the breaks in sub-samples $\hat{\mathcal{T}}_j^*$ - see Boldea, Hall, and Han (2012) - one obtains consistent estimates of the non-common break fractions in (18). Denote their break-point counterparts by $\hat{T}_n, (n = 1, 2, \ldots, N)$, with $N \leq m$. These are the breaks that are idiosyncratic to the second-stage.

- Let $\hat{T}_0 = 0$ and $\hat{T}_{N+1} = T$, to include sample ends. Obtain the union of sample end-points and the breaks in the first-stage and second-stage, ordered, as $\mathcal{B} = \{\hat{T}_0, \ldots, \hat{T}_{N+1}\} \cup \{\hat{T}_1^*, \ldots, \hat{T}_{\hat{m}^*}^*\}$. Thus, $\mathcal{B}$ contains all the breaks in the first- and second-stage equation (18) and (19).

Thus, via this strategy, the researcher knows the idiosyncratic breaks to the second-stage, and all the breaks in the first-stage. However, for practical purposes, one needs to know which breaks are common to the two stages. This is not only important for correct estimation of the sub-samples in the second-stage, it is also of interest to practitioners. For example, in the estimation of a hybrid NKPC, with measured expectations, Boldea, Hall, and Han (2012) show that a break at the end of 1980 occurred in the modeled inflation expectations, but that break did not further occur in the NKPC itself once the change in expectations was taken into account. The test they use to detect common breaks is a usual Wald test for a known break-point, and is defined below.

(v-ii) Tests for breaks that are common to both stages

To describe these tests, let the ordered breaks in $\mathcal{B}$ be $\tilde{T}_1, \tilde{T}_2, \ldots, \tilde{T}_H$, with $H = \hat{m}^* + N$. Then for any $s = 1, 2, \ldots, H$ such that $\hat{T}_j^* = \tilde{T}_s$, take the smallest sub-sample encompassing it, $\{\tilde{T}_{s-1} + 1, \ldots, \tilde{T}_{s+1}\}$, and calculate the 2SLS estimators $\hat{\theta}_s$ and $\hat{\theta}_{s+1}$, based on sub-samples $\mathcal{B}_s = \{\tilde{T}_{s-1} + 1, \ldots, \tilde{T}_s\}$, respectively $\mathcal{B}_{s+1} = \{\tilde{T}_s + 1, \ldots, \tilde{T}_{s+1}\}$. Treat the end-points $\tilde{T}_{s-1}$ and $\tilde{T}_{s+1}$, and $\hat{T}_j^*$ as known. Then the Wald test for a common break $\tilde{T}_s$, to both equations (18) and (19), is:

$$\mathcal{W}_T = T(\hat{\theta}_s - \hat{\theta}_{s+1})'[V_{s,T} + V_{s+1,T}]^{-1}(\hat{\theta}_s - \hat{\theta}_{s+1}), \tag{20}$$

where $V_{i,T}$ are consistent estimates of the asymptotic variances $V_i = \lim_{T\to\infty} Var$ $[T^{1/2}(\hat{\theta}_i - \theta_i)]$, for $i = s, s+1$. Because we test in the second-stage for a break pre-estimated from another equation, the first-stage, the distribution is the same as if the break-point $\hat{T}_s$ were known.

Proposition 6 *Under certain regularity conditions*[17]*, under the null hypothesis of no common break in the sub-sample tested,*

$$\mathcal{W}_T \xrightarrow{d} \chi^2_1 \tag{21}$$

Thus, all the breaks in the equation of interest (18) can be retrieved via the sequential strategy in (v) - or (iv) if no breaks are found in the first-stage regression.

This procedure was defined for one endogenous regressor, but can be generalized to p multiple endogenous regressors X_t. Intuitively, one just needs to consistently predict the endogenous regressors from the first stage, so the OLS sequential strategy in (i)-(iii) can be applied to the first stage equation pertaining to each endogenous regressor separately.[18] If the second-stage equation also has exogenous regressors a_t, then the tests in (iv)-(v) remain valid, with $\hat{x}_t$ replaced by $(\hat{X}_t', a_t')'$. The optimality of these procedures is, as for OLS methods, unclear, but all the tests reject with probability one for large T, regardless of whether the breaks are small or large.

In nonlinear models, a similar sequential procedure as in (i)-(iii) is available for models that can be estimated via nonlinear least squares - see Boldea and Hall (2010). Given a known parametric regression function with multiple parameter changes, the tests in (i)-(iii) remain valid, and so do their distributions, as long as the OLS estimators in (i)-(iii) are replaced by their NLS counterparts.

Even though a procedure for testing for multiple breaks in linear models estimated via GMM is not known, its 2SLS counterpart can be used, and as a byproduct, one obtains consistent break-points as well as parameter estimates. Thanks to the dynamic programming algorithm introduced in Bai and Perron (2003), the computational burden for a sample size of T is less than $T(T+1)/2$ operations independent of the number of breaks. GAUSS code for testing for multiple breaks can be found at *http : //people.bu.edu/perron/code.html* for OLS methods.

The testing procedures above are all designed for multiple breaks, when the parameter changes infrequently and permanently from one value to another at a few locations in the sample. Other types of parameter instability are summarized in the next section.

[17] See Hall, Han, and Boldea (2012).

[18] For estimating the breaks by considering a multivariate first-stage for all endogenous regressors jointly, a comprehensive procedure can be found in Qu and Perron (2007).

4 Testing for other types of parameter change

Break-points are by default exogenous to the model, since time is an exogenous quantity. When parameter changes are believed to be driven by some observed variables that indicate the state of the business cycle or some other important economic indicators, researchers resort to other types of parameter change models, called threshold and smooth transition models.

Threshold and smooth transition models have been used to model GDP growth, unemployment, interest rates, prices, stock returns and exchange rates - for a review of the empirical and theoretical econometrics literature, see van Dijk, Teräsvirta, and Franses (2002) and Hansen (2011).

The threshold model, introduced by Howell Tong[19], resembles in many ways the break-point model. To see the connection, for simplicity we revert to an exposition with one break. Thus, consider the model (4) but $m = 1$, re-written in the following way:

$$y_t \;=\; x_t'\theta_t + u_t \text{ with } \theta_t = \theta_1 + (\theta_2 - \theta_1)\mathbf{1}\{t \geq T_1\} \tag{22}$$

where $\mathbf{1}$ is the indicator function. If instead, θ_t is defined as:

$$y_t \;=\; x_t'\theta_t + u_t \text{ with } \theta_t = \theta_1 + (\theta_2 - \theta_1)\mathbf{1}\{q_t \geq c\}, \tag{23}$$

where y_t is a scalar dependent variables, c is an unknown parameter, to be estimated, and x_t and q_t are exogenous observed regressors, uncorrelated with u_t, one obtains the *threshold model.* If x_t contains lags of y_t, this model is known as the *threshold autoregressive (TAR) model.*

In this model, the parameter change is driven by the observed variable q_t, called a state variable, and c denotes the threshold above which the parameters of the model change from θ_1 to θ_2. If one orders the data (y_t, x_t) on the values of q_t, say in ascending order, one obtains two sub-samples, one for which the true parameter is θ_1, and another for which the true parameter is θ_2. These two sub-samples resemble the break-point sub-samples; the break-point here is the point in the newly ordered sample where q_t changes from a value below c to a value above c. Hansen (1997b) shows that by re-ordering of the data as described above, a $sup\mathcal{W}_T$ of the type described in the previous section can be constructed, and its asymptotic distribution for one threshold is the same as in Proposition 3 for $k = 1$. This procedure is extended to estimating multiple thresholds in Gonzalo and Pitarakis (2002).

When some of the regressors in x_t are endogenous, but q_t is exogenous, for one threshold model, Caner and Hansen (2004) propose the *sup*-type Wald test constructed with a 2SLS estimator of the threshold c and GMM estimators of the other parameters. Its asymptotic distribution is computed by simulation in Caner and Hansen (2004). For endogenous q_t, no tests are known so far, although estimation of models with one endogenous threshold can be done via 2SLS estimation with bias correction - see Kourtellos, Stengos, and Tan (2008).

[19]For an early review of the threshold literature in statistics, see Tong (1983).

An equally influential model in empirical macroeconomics is the *smooth transition autoregressive (STAR) model*, introduced by Teräsvirta (1994). This model is mostly suitable for policy functions such as the interest rate functions, where the parameter is not changing in a sudden fashion from θ_1 to θ_2, but in a smooth way. The smoothness is imputed by replacing the indicator function with a smooth function, with values in the interval $(0, 1)$, meaning that true parameters at each point in time no longer take two values, θ_1 and θ_2, but are most of the time in-between these values. We present here the simplest STAR model, called the logistic STAR, where the indicator function is replaced by the smooth logistic function $f(\cdot, \cdot, \cdot)$, defined below:

$$y_t = x_t'\theta_t + u_t \text{ with } \theta_t = \theta_1 + (\theta_2 - \theta_1) f(q_t, \gamma, c) \tag{24}$$

$$f(q_t, \gamma, c) = \frac{exp\{\gamma(q_t - c)\}}{1 + exp\{\gamma(q_t - c)\}} \tag{25}$$

Here q_t and x_t are observed and uncorrelated with the errors u_t, and may contain lags of y_t, γ is an unknown smoothness parameter and c is the unknown 'threshold' parameter. The latter two need to be estimated. When γ is large, the transition is faster as the transition function moves faster from 0 to 1 when q_t is above c; when γ is small, the transition is slower.

This model is different than the threshold or break model in the sense that it can be written as a smooth nonlinear regression. Unfortunately, the usual nonlinear regression Wald tests do not directly apply. The reason is because, under the null hypothesis of $\theta_1 = \theta_2$, or no nonlinearity, the parameters γ, c are not identified, since any value of these parameters will render the same linear regression function. This implies that we may have a similar problem to the break-point or threshold model, of unidentified parameters under the null hypothesis. However, because the regression function in (23) is smooth under the alternative, one can write a third-order Taylor expansion of $f(q_t, \cdot, c)$ around $\gamma = 0$. This yields a model that has the parameters γ, c also identified under the null hypothesis, and Eitrheim and Teräsvirta (1996) show that the resulting model can be tested for no nonlinearity via a usual LM test. If one doesn't use this expansion, the optimality properties of a *Sup*-type LR test over all possible values of γ, c, are studied in Andrews, Cheng, and Guggenberger (2011), who show that this test may not have the right size.

STAR models have recently been shown to be estimable via GMM in the presence of endogenous regressors using the usual GMM asymptotic theory - see Areosa, McAleer, and Medeiros (2011). Thus, we conjecture that the tests discussed here for STAR models can be used in their equivalent GMM form with no further complications.

So far, we have discussed models for parameter change driven by an observed variable. However, one may think that the behavior of macroeconomic variables changes with an unobserved variable such as the state of the business cycle. Such models are called regime-switching models (or Markov-switching under certain regularity conditions) and were introduced in econometrics by Hamilton (1989). To write down a Markov-switching model, let s_t be an unobserved state variable, with values zero or one (i.e. for recession or expansion). Then:

$$y_t = x_t'\theta_t + u_t \text{ with } \theta_t = \theta_1 + (\theta_2 - \theta_1)s_t \tag{26}$$

$$P(s_t = 1|s_{t-1} = 1, \mathcal{I}_t) = p_1; P(s_t = 0|s_{t-1} = 0, \mathcal{I}_t) = p_2 \tag{27}$$

Here, x_t usually includes some lags of y_t, and is uncorrelated with u_t. Also, $s_t \in \{0, 1\}$, and the probabilities of being in a certain state are entirely determined by the previous state and the information set at time t, $\mathcal{I}_t$, which includes x_t and all its previous values.

Suppose one wants to test whether the parameter change $\theta_2 - \theta_1$ is zero or not. In this case, p and q are the parameters that not identified under the null hypothesis, but there are other complications related to the cases of p or q being close to zero or one. This presence implies that we cannot use Taylor expansions as in the STAR example to test for parameter change. Hansen (1992) proposes an upper bound for a *sup*-type LR test, but this bound depends on the data and is often burdensome to compute. Garcia (1998) proposes to restrict testing to cases where p, q are bounded away from $0, 1$, and use the *sup*-LR test, where the supremum is taken over (p, q), and gives the asymptotic distribution of the test. However, because the framework he uses to justify his test, taken from Andrews and Ploberger (1994), does not apply to Markov-switching models, this test may not have optimal power. To that end, we recommend the test by Carrasco, Hu, and Ploberger (2009), an information-matrix type LM test that is shown to have certain desired optimality properties.

When x_t is endogenous, Kim (2004) and Kim (2009) show that either a bias-corrected maximum-likelihood estimation or a two-step maximum likelihood ignoring the bias can be used the estimate the Markov-switching model with two regimes. Similarly, when s_t is endogenous but x_t are exogenous, Kim, Piger, and Startz (2008) propose a bias-corrected filter to estimate the model. As for STAR, we conjecture that tests can be constructed based on these inference procedures, adapted from the Markov switching tests for exogenous regressors.

There are many other types of parameter change, but the main types are summarized here and they all have their merits for the applied researcher.

5 Testing for other types of structural instability

So far, the focus of this chapter has been on testing for parameter change; however this is not the only scenario that can lead to structural instability. In this section, we explore tests for other forms of structural instability that have been developed within the GMM framework. To this end, we return to the framework in Section 2.

A natural alternative to H_0 in (1) that captures the notion of structural instability at a fixed break-point $T_1 = [T\lambda]$ is

$$H_1'(\lambda): \begin{array}{rcll} E[f(v_t,\theta_0)] & = & 0, & \text{for } t = 1,2,\ldots,T_1, \\ E[f(v_t,\theta_0)] & \neq & 0, & \text{for } t = T_1+1,\ldots,T. \end{array}$$

Under $H_1'(\lambda)$, the population moment holds at θ_0 before the break-point, but fails to hold after. Before proceeding further, we note that all that follows applies equally if the population moment condition holds at θ_0 after but not before the break-point.

If $q = p$ - and so there are the same number of moment conditions as parameters - then it can be shown that $H_1(\lambda)$ and $H_1'(\lambda)$ are equivalent.[20] However, if $q > p$ - and so there are more moment conditions than parameters - then this equivalence does not hold. Exploiting the decomposition of the population moment condition inherent in GMM estimation,[21] Hall and Sen (1999) show that $H_1'(\lambda)$ can be decomposed into two parts: structural instability in the identifying restrictions at $[T\lambda]$ and structural instability in the overidentifying restrictions at $[T\lambda]$. Instability of the identifying restrictions is equivalent to $H_1(\lambda)$ that is, to parameter change at $T\lambda$. Instability of the overidentifying restrictions means that some aspect of the model beyond the parameters alone has changed at $[T\lambda]$. Rather than specifying this alternative mathematically, we present an intuitive explanation. To this end, consider Hansen and Singleton's (1982) consumption based asset pricing in which a representative agent makes consumption and investment decisions to to maximize discounted expected lifetime utility based on a utility function,

$$U_t(c_t) = \frac{c_t^{\gamma} - 1}{\gamma}.$$

The parameters of this model are γ, with $1-\gamma$ being the coefficient of relative risk aversion, and β, the discount factor; so, in our notation, $\theta = (\gamma, \beta)'$. It is customary to estimate θ_0 via GMM based on $E[u_t(\theta_0)z_t] = 0$ where $u_t(\theta_0)$ is the so-called Euler residual derived from the underlying economic model, and z_t is a vector of variables contained in the representative agent's information set. Within this setting, $H_1(\lambda)$ implies that the parameters have changed but the functional form of $u_t(\theta)$ has stayed the same: this state of the world occurs if the functional form of the agent's utility function is the same before and after the break-point but either his/her coefficient of relative risk aversion or his/her discount factor changes. Instability of the overidentifying restrictions implies that the structural change involves more than the parameters: this state of the world would occur if the functional form of the agent's utility function changes at the break-point.

Since instability of the identifying and overidentifying restrictions have different implications for the underlying model, Hall and Sen (1999) argue that it

[20] This result presumes certain (relatively weak) regularity conditions hold and can be established via a similar argument to Hall and Inoue (2003)[Proposition 1].

[21] For example, see Chapter 14.

is advantageous to test $H_0(\lambda)$ against each separately. To test against instability of the identifying restrictions, we can use the statistics for testing against $H_1(\lambda)$ described in Section 2. To test against instability of the overidentifying restrictions, Hall and Sen (1999) propose the use of the following statistic,

$$\mathcal{O}_T(\lambda) \;=\; \mathcal{O}_{1,T}(\lambda) \;+\; \mathcal{O}_{2,T}(\lambda)$$

where $\mathcal{O}_{i,T}(\lambda)$ is the overidentifying restrictions tests based on the sub–sample $\mathcal{T}_i$ for $i = 1, 2$.[22] The following proposition gives the limiting distribution of $\mathcal{O}_T(\lambda)$.

Proposition 7 *If certain regularity conditions hold then under H_0, then we have:* $\mathcal{O}_T(\lambda) \xrightarrow{d} \chi_{2(q-p)}$.

Hall and Sen (1999) show that under H_0, $\mathcal{O}_T(\lambda)$ is asymptotically independent of the statistics, such as $\mathcal{W}_T(\lambda)$, used to test parameter change. They further show that under local alternatives each of $\mathcal{W}_T(\lambda)$ and $\mathcal{O}_T(\lambda)$ has power against its own specific alternative but none against the alternative of the other test. These properties underscore the notion that the two statistics are testing different aspects of instability and provides some guidance on the interpretation of significant statistics, although the local nature of the results needs to be kept in mind.[23]

Hall and Sen (1999) also propose the following statistics for testing for instability of the overidentifying restrictions at an unknown break-point,[24]

$$\begin{aligned}
Sup\mathcal{O}_T &= \sup_{i\in T_b} \{\, \mathcal{O}_T(i/T) \,\} \\
Av\mathcal{O}_T &= d_b^{-1} \sum_{i=[\lambda_l T]}^{[\lambda_u T]} \mathcal{O}_T(i/T) \\
Exp\mathcal{O}_T &= log \left\{ d_b^{-1} \sum_{i=[\lambda_l T]}^{[\lambda_u T]} exp[0.5\mathcal{O}_T(i/T)] \right\}
\end{aligned}$$

The limiting distribution of these statistics is as follows.

[22] See Chapter 14 for a description of the overidentifying restrictions test.

[23] To elaborate, $\mathcal{W}_T(\lambda)$ significant but $\mathcal{O}_T(\lambda)$ insignificant is consistent with instability confined to the parameters alone; $\mathcal{O}_T(\lambda)$ significant is consistent with more general forms of instability. However, important caveats are the local nature of these results and concerns about the interpretation of fixed break-point tests discussed in Section 2.

[24] Although the functionals are the same as the parameter change tests, it has proved impossible to date to deduce any optimality properties for the versions based on $O_T(\lambda)$ due to the nature of the alternative in this case; see Hall (2005)[p.182] for further discussion.

Proposition 8 *If certain regularity conditions hold then under H_0, then we have: $Sup\mathcal{O}_T \Rightarrow Sup_{\lambda\in\Lambda}\mathcal{O}(\lambda)$, $Av\mathcal{O}_T \Rightarrow \int_\Lambda \mathcal{O}(\lambda)dJ(\lambda)$, and $Exp\mathcal{O}_T \Rightarrow log\left[\int_\Lambda exp\{0.5\,\mathcal{O}(\lambda)\}dJ(\lambda)\right]$ where $\mathcal{O}(\lambda) = \frac{1}{\lambda}B'_{q-p}B_{q-p} + \frac{1}{1-\lambda}[B_{q-p}(1) - B_{q-p}(\lambda)]'[B_{q-p}(1) - B_{q-p}(\lambda)]$ and $B_{q-p}(\lambda)$ denotes a $(q-p)\times 1$ Brownian motion on $[0,1]$.*

The limiting distributions in Proposition 8 are non-standard but depend only on $q-p$, the number of overidentifying restrictions. The percentiles of these limiting distributions are reported in Hall and Sen (1999). Sen and Hall (1999) report response surfaces which can be used to calculate approximate p–values for all three versions of these tests.

6 Conclusions

In this chapter, we describe various structural stability tests that are best suited for empirical analysis in macroeconomics. We discuss tests for parameter breaks, other types or parameter changes and also model instability. Most of our discussion is focused around Wald-type tests, that are robust to heteroskedasticity and autocorrelation. The tests we described do not cover all possible scenarios, such as the presence of unit roots, long memory, cointegrating relationships. Instead, we strive to provide practitioners with a broad set of tools that cover most cases of interest - excluding the above - in empirical macroeconometrics. We emphasize on structural stability tests for models with exogenous and endogenous regressors and the differences between the two.

The tools we summarize here can be readily applied to a wide range of macroeconomic and monetary policy models, contributing to important macroeconomic debates such as whether the New Keynesian Phillips curve has become less predictable, or whether the Great Moderation is due to good monetary policy or good luck - see Lubik and Surico (2010).

References

Alogoskoufis, G., and Smith, R. (1991). 'The Phillips curve, the persistence of inflation, and the Lucas critique: Evidence from exchange-rate regimes', *American Economic Review*, pp. 1254–1275.

Andrews, D., Cheng, X., and Guggenberger, P. (2011). 'Generic results for establishing the asymptotic size of confidence sets and tests', *Cowles Foundation Discussion Paper No. 1813*.

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

——— (2003). 'Tests for parameter instability and structural change with unknown change point: a corrigendum', *Econometrica*, 71: 395–398.

Andrews, D. W. K., and Fair, R. (1988). 'Inference in econometric models with structural change', *Review of Economic Studies*, 55: 615–640.

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

Areosa, W., McAleer, M., and Medeiros, M. (2011). 'Moment-based estimation of smooth transition regression models with endogenous variables', *Journal of Econometrics*, 165: 100–111.

Bai, J. (2009). 'Likelihood ratio tests for multiple structural changes', *Journal of Econometrics*, 91: 299–323.

Bai, J., and Perron, P. (1998). 'Estimating and testing linear models with multiple structural changes', *Econometrica*, 66: 47–78.

——— (2003). 'Computation and Analysis of Multiple Structural Change Models', *Journal of Applied Econometrics*, 18: 1–23.

Ben-David, D., Lumsdaine, R., and Papell, D. (2003). 'Unit roots, postwar slowdowns and long-run growth: Evidence from two structural breaks', *Empirical Economics*, 28(2): 303–319.

Benati, L., and Surico, P. (2008). 'Evolving US monetary policy and the decline of inflation predictability', *Journal of the European Economic Association*, 6(2-3): 634–646.

Boivin, J., and Giannoni, M. (2006). 'Has monetary policy become more effective?', *The Review of Economics and Statistics*, 88(3): 445–462.

Boldea, O., Cornea, A., and Hall, A. (2011). 'Bootstrapping multiple structural change tests in models estimated via 2SLS', *Working Paper*.

Boldea, O., and Hall, A. R. (2010). 'Estimation and Inference in unstable nonlinear least squares models', *Working Paper, University of Manchester.*

Boldea, O., Hall, A. R., and Han, S. (2012). 'Asymptotic distribution theory for break point estimators in models estimated via 2SLS', *Econometric Reviews*, 33: 1–33.

Caner, M. (2011). 'Pivotal structural change tests in linear simultaneous equations with weak identification', *Econometric Theory*, 1(1): 1–14.

Caner, M., and Hansen, B. (2004). 'Instrumental Variable Estimation of a Threshold Model', *Econometric Theory*, 20: 813–843.

Carrasco, M., Hu, L., and Ploberger, W. (2009). 'Optimal test for Markov Switching parameters', *Working Paper, University of Montreal, Canada.*

Clarida, R., Gali, J., and Gertler, M. (2000). 'Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory', *Quarterly Journal of Economics*, 115: 147–180.

Cogley, T., and Sargent, T. (2001). 'Evolving post-world war II US inflation dynamics', in *NBER Macroeconomics Annual.* MIT Press.

Eitrheim, Ø., and Teräsvirta, T. (1996). 'Testing the adequacy of the smooth transition autoregressive models', *Journal of Econometrics*, 74: 59–75.

Estrella, A., and Fuhrer, C. (2003). 'Monetary policy shifts and the stability of monetary policy models', *Review of Economics and Statistics*, 85(1): 94–104.

Garcia, R. (1998). 'Asymptotic null distribution of the likelihood ratio test in Markov Switching models', *International Economic Review*, 39: 763–788.

Gonzalo, J., and Pitarakis, J.-Y. (2002). 'Estimation and model selection based inference in single and multiple threshold models', *Journal of Econometrics*, 110: 319–352.

Hall, A. R. (2005). *Generalized Method of Moments.* Oxford University Press, Oxford, U.K.

——— (2011). 'Generalized Method of Moments', Discussion paper, prepared for in inclusion in Handbook of Research Methods and Applications on Empirical Macroeconomics, N. Hashimzade and M. Thornton ed.s.

Hall, A. R., Han, S., and Boldea, O. (2012). 'Inference regarding multiple structural changes in linear models with endogenous regressors', *Journal of Econometrics*, forthcoming.

Hall, A. R., and Inoue, A. (2003). 'The large sample behaviour of the Generalized Method of Moments estimator in misspecified models', *Journal of Econometrics*, 114: 361–394.

Hall, A. R., and Sakkas, N. (2011). 'Approximate p-values of Certain Tests Involving Hypotheses about Multiple Breaks', *Journal of Econometric Methods*, Accepted for publication.

Hall, A. R., and Sen, A. (1999). 'Structural stability testing in models estimated by Generalized Method of Moments', *Journal of Business and Economic Statistics*, 17: 335–348.

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

Hansen, B. E. (1992). 'The Likelihood Ratio test under nonstandard conditions: testing the Markov Switching model of GNP', *Journal of Applied Econometrics*, 7: S61–S82.

——— (1997a). 'Approximate asymptotic p–values for structural change tests', *Journal of Business and Economic Statistics*, 15: 60–67.

——— (1997b). 'Inference in TAR models', *Studies in Nonlinear Dynamics and Econometrics*, 2.

——— (2000). 'Testing for structural change in conditional models', *Journal of Econometrics*, 97(1): 93–115.

——— (2011). 'Threshold autoregression in economics', *Statistics and its interface*, 4: 123–127.

Hansen, L. P., and Singleton, K. J. (1982). 'Generalized instrumental variables estimation of nonlinear rational expectations models', *Econometrica*, 50: 1269–1286.

Kang, K., Kim, C., and Morley, J. (2009). 'Changes in US inflation persistence', *Studies in Nonlinear Dynamics & Econometrics*, 13(4).

Kim, C.-J. (2004). 'Markov-switching models with endogenous explanatory variables', *Journal of Econometrics*, 122: 127–136.

——— (2009). 'Markov-switching models with endogenous explanatory variables II: A two-step MLE procedure', *Journal of Econometrics*, 148: 46–55.

Kim, C.-J., Piger, J., and Startz, R. (2008). 'Estimation of Markov regime-switching regression models with endogenous switching', *Journal of Econometrics*, 143: 263–273.

Kim, D., and Perron, P. (2009). 'Assessing the relative power of structural break tests using a framework based on the approximate Bahadur slope', *Journal of Econometrics*, 149: 26–51.

Kourtellos, A., Stengos, T., and Tan, C. (2008). 'THRET: Threshold Regression with Endogenous Threshold Variables', *University of Cyprus Discussion Paper No. 3.*

Liu, Z., Waggoner, D., and Zha, T. (2009). 'Asymmetric expectation effects of regime shifts in monetary policy', *Review of Economic Dynamics*, 12(2): 284–303.

Lubik, T., and Surico, P. (2010). 'The Lucas critique and the stability of empirical models', *Journal of Applied Econometrics*, 25(1): 177–194.

Lucas, R. (1976). 'Econometric policy evaluation: a critique', in K. Brunner and A. Melzer (eds.), *The Phillips Curve and Labor Markets*, vol. 1 of *Carnegie-Rochester Conference Series on Public Policy*, pp. 19–46.

Morgan, M. (1990). *The History of Econometric Ideas*. Cambridge University Press, New York, NY, U. S. A.

Papell, D., Murray, C., and Ghiblawi, H. (2000). 'The structure of unemployment', *Review of Economics and Statistics*, 82(2): 309–315.

Perron, P. (1997). 'Further evidence on breaking trend functions in macroeconomic variables', *Journal of econometrics*, 80(2): 355–385.

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

Rossi, B. (2006). 'Are exchange rates really random walks? Some evidence robust to parameter instability', *Macroeconomic Dynamics*, 10(01): 20–38.

Sen, A. (1997). 'New tests of structural stability and applications to consumption based asset pricing models', Ph.D. thesis, Department of Economics, North Carolina State University, Raleigh, NC.

Sen, A., and Hall, A. R. (1999). 'Two further aspects of some new tests for structural stability', *Structural Change and Economic Dynamics*, 10: 431–443.

Sowell, F. (1996). 'Optimal tests of parameter variation in the Generalized Method of Moments framework', *Econometrica*, 64: 1085–1108.

Surico, P. (2007). 'The Fed's monetary policy rule and US inflation: The case of asymmetric preferences', *Journal of Economic Dynamics and Control*, 31(1): 305–324.

Teräsvirta, T. (1994). 'Specification, estimation, and evaluation of smooth transition autoregressive models', *Journal of the American Statistical Association*, 89: 208–218.

Tinbergen, J. (1939). *Statistical testing of business-cycle theoreis, Volumes I & II*. League of Nations, Gevea, switzerland.

Tong, H. (1983). *Threshold models in non-linear time series analysis*, vol. 21 of *Lecture Notes in Statistics*. Springer, Berlin, Germany.

van Dijk, D., Teräsvirta, T., and Franses, P. (2002). 'Smooth transition autoregressive models: a survey or recent developments', *Econometric Reviews*, 21: 1–47.

Weber, C. (1995). 'Cyclical output, cyclical unemployment, and Okun's coefficient: a new approach', *Journal of Applied Econometrics*, 10(4): 433–445.

Zhang, C., Osborn, D., and Kim, D. (2008). 'The new Keynesian Phillips curve: from sticky inflation to sticky prices', *Journal of Money, Credit and Banking*, 40: 667–699.