Joint Detection of Structural Change and Nonstationarity in Autoregressions

Jean-Yves Pitarakis
University of Southampton
Economics Division
Southampton SO17 1BJ, U.K

February 14, 2011

Abstract

In this paper we develop a test of the joint null hypothesis of parameter stability and a unit root within an ADF style autoregressive specification whose entire parameter structure is potentially subject to a structural break at an unknown time period. The maintained underlying null model is a linear autoregression with a unit root, stationary regressors and a constant term. We derive the limiting distribution of a Supremum Wald type test statistic and show that it can be expressed as the sum of two components each corresponding to the nonstationary and stationary regressors respectively. As a byproduct we also obtain the limiting behaviour of a related Wald statistic designed to solely test the null of parameter stability in an environment with a unit root. These distributions are free of nuisance parameters and easily tabulated. The finite sample properties of our tests are subsequently assessed through a series of simulations.

Key Words: Structural Breaks, Unit Roots, Nonlinear Dynamics.

[1]Financial support from the ESRC is gratefully acknowledged. Address for Correspondence: Jean-Yves Pitarakis, University of Southampton, School of Social Sciences, Economics Division, Southampton, SO17 1BJ, United-Kingdom. Email: j.pitarakis@soton.ac.uk

1 Introduction

A vast body of research in the recent time series econometrics literature has explored the interactions between nonlinear dynamics and unit root type of nonstationarities. Although initially nonlinearities and nonstationarities were often treated as separate and sometimes mutually exclusive phenomena the development of new functional central limit theory amongst other technical tools has led to a growing body of research dealing with models in which both features could coexist. Under structural break type of nonlinearities for instance and starting with the early work of Perron (1989) there has been a vast literature on designing unit root tests that allowed for the presence of breaks in the underlying deterministic trend function of a series. One motivation for this line of research was the observation that the omission or misspecification of such trend breaks could lead to misleading inferences about the presence of unit roots. Important contributions in this area include Zivot and Andrews (1992), Banerjee, Lumsdaine and Stock (1992) and more recently Kim and Perron (2009), Harris, Harvey, Leybourne and Taylor (2009) amongst numerous others. The complications induced by the coexistence of structural breaks and unit roots have also triggered an interesting research agenda that instead focused on the impact of unit roots on Chow type parameter stability tests and documented a spurious break phenomenon (see Bai (1998)) whereby ignoring the presence of a unit root in an otherwise linear model was shown to frequently lead to the detection of spurious break points. Despite the voluminous literature that explored these issues numerous open questions on the impact of nonstationarity on tests for structural breaks still remain.

In this paper our goal is to explore the joint interaction of structural change and unit roots by allowing the parameters of both the deterministic and stochastic components of an augmented Dickey-Fuller (ADF) type autoregressive model to be subject to a structural break. Unlike the existing literature that has mainly seeked to robustify unit root inferences to trend breaks and other related features we instead concentrate on detecting the presence of parameter instability and nonstationarity in an ADF specification whose autoregressive parameters may also be subject to structural breaks. More specifically, we are interested in exploring the properties of a Wald type test statistic designed to test the joint hypothesis of parameter stability and a unit root within an ADF style autoregression. We view our test as a useful and practical diagnostic tool for further enhancing the existing apparatus on structural break and unit root testing. Subject to some confidence level for instance, a non rejection of our joint null of a unit root and parameter stability may preclude the need to undertake further break point or related analyses. In addition, and unlike traditional unit root tests our new test is also shown to have a strong ability to detect switches from a unit root to a stationary regime and vice versa.

As a byproduct of the above objectives we also derive the properties of a related Wald statistic whose sole purpose is to test the constancy of all the parameters characterising an ADF style autoregression when a unit root is imposed in the underlying model. This latter test statistic will help highlight the consequences of ignoring the presence of a unit root on commonly used tests for structural breaks. Finally, we also view the motivation of this paper as following closely Caner and Hansen (2001) where the authors explored similar issues in models characterised by threshold effects as opposed to the structural break setting considered here. This comparison allows us to make interesting parallels between the two very different ways of capturing change.

The plan of the paper is as follows. Section 2 presents our operating model and motivates the hypotheses of interest. Section 3 develops the large sample theory of our Wald type test statistics. Section 4 provides numerical simulations and Section 6 concludes. All proofs are relegated to the appendix.

2 The Model and Hypotheses

Our operating model is given by the familiar ADF specification with all the parameters of its deterministic and stochastic components allowed to switch at some unknown time period k. Specifically, we consider

$$\Delta y_t = \begin{cases} \alpha_1 + \beta_1 t + \rho_1 y_{t-1} + \sum_{j=1}^{p} \gamma_{1j} \Delta y_{t-j} + e_t & t \leq k \\ \alpha_2 + \beta_2 t + \rho_2 y_{t-1} + \sum_{j=1}^{p} \gamma_{2j} \Delta y_{t-j} + e_t & t > k \end{cases} \tag{1}$$

with e_t denoting an iid disturbance. It is also convenient to reformulate (1) in matrix form as

$$\Delta Y \quad = \quad X_1 \theta_1 + X_2 \theta_2 + e \tag{2}$$

with $\Delta Y = (\Delta y_1, \ldots, \Delta y_T)'$. Letting $r_t = (1\ t)'$ and $z_{t-1} = (\Delta y_{t-1}, \ldots, \Delta y_{t-p})'$, X_1 above stacks the elements of $(r_t'\ y_{t-1}\ z_t')I(t \leq k)$, X_2 those of $(r_t'\ y_{t-1}\ z_t')I(t > k)$ and $\theta_i = (\alpha_i\ \beta_i\ \rho_i\ \gamma_{i1}\ \ldots \gamma_{ip})'$ for $i = 1, 2$. Throughout this paper k will denote the unknown breakpoint location and for later use we also introduce the break fraction $\pi = \lim_{T \to \infty} k/T$ with $\pi \in [\underline{\pi}, \overline{\pi}] \subset (0, 1)$. For notational simplicity we will also refer to the two indicator functions as $I_1 \equiv I(t \leq k)$ and $I_2 \equiv I(t > k)$. Letting $X = X_1 + X_2$ denote the matrix that stacks the elements $(r_t'\ y_{t-1}\ z_{t-1}')$ of the linear model it is also convenient to reparameterise (2) as

$$\Delta Y \quad = \quad X \theta_2 + X_1 \Psi + e \tag{3}$$

with $\Psi \equiv (\theta_1 - \theta_2)$.

Our main concern is to develop a test of the joint null of a unit root and the absence of a structural break in all the ADF parameters. We write this hypothesis as $H_0^A : \Psi = 0, \rho_1 = 0$. A non rejection of this null would indicate support for the presence of a unit root in y_t together with the suitability of a linear autoregressive specification while precluding the need to explore further the potential presence of breakpoints in some or all of the ADF parameters. In this sense, we view the implementation of a test such as $H_0^A : \Psi = 0, \rho_1 = 0$ as a useful diagnostic tool. Furthermore and as demonstrated below we expect our test to display a strong ability to detect scenarios where ρ_i switches from zero to a stationary region such as $\{\rho_1 = 0, \rho_2 < 0\}$ or $\{\rho_1 = 0, \rho_2 < 0\}$. As a byproduct of our theory underlying $H_0^A : \Psi = 0, \rho_1 = 0$ we also obtain the limiting distribution of a Wald type test statistic for testing the null of parameter constancy formulated as $H_0^B : \Psi = 0$. An important goal here is to use our distributional theory surrounding $H_0^B : \Psi = 0$ to formally highlight the dangers of ignoring the presence of a unit root when implementing breakpoint tests via standard methods (e.g. following the asymptotic theory developed in Andrews (1992)). At this stage it is also important to point out that throughout this paper and as in Caner and Hansen (2001) our maintained model under our null hypotheses is given by $\Delta y_t = \alpha + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + e_t$ and rules out the presence of any deterministic trend components.

Viewing the model in (3) as our most general specification it is easy to note that its corresponding sum of squared residuals, say SSR_{MG}, can be written as $SSR_{MG} = \Delta Y' M_{X,X_1} \Delta Y$ with $M_{X,X_1} = M_X - M_X X_1 (X_1' M_X X_1)^{-1} X_1' M_X$ and $M_X = I - X(X'X)^{-1}X'$. Letting W stack the regressors of the model restricted by $H_0^A : \Psi = 0, \rho_1 = 0$ we have $SSR_A = \Delta Y' M_W \Delta Y$ so that $W_T^A(k) = [\Delta Y' M_W \Delta Y - \Delta Y' M_{X,X_1} \Delta Y]/\hat{\sigma}_e^2$. Similarly, for the model restricted by $H_0^B : \Psi = 0$ we have $SSR_B = \Delta Y' M_X \Delta Y$ so that the standard Wald statistic for testing $H_0^B : \Psi = 0$ and a given k can now be formulated as $W_T^B(k) = [\Delta Y' M_X \Delta Y - \Delta Y' M_{X,X_1} \Delta Y]/\hat{\sigma}_e^2$ with $\hat{\sigma}_e^2$ denoting the residual variance from (3). In practice since the break parameter is unidentified under the null hypothesis inferences are conducted using the well known supremum versions of $W_T^A(k)$ and $W_T^B(k)$. Following common practice in the literature we trim a percentage of the top and bottom of the sample by setting $[k_1, k_2] = [[T\pi_1], [T\pi_2]]$ and using $\pi_2 = 1 - \pi_1$ with $\pi_1 = 0.10$ so that our test statistics are now given by $SupWald^A \equiv \sup_{\pi \in [\pi_1, \pi_2]} W_T^A(\pi)$ and $SupWald^B \equiv \sup_{\pi \in [\pi_1, \pi_2]} W_T^B(\pi)$.

3 Large Sample Inference

In what follows we will operate under assumptions that are similar to those maintained in Caner and Hansen (2001). Throughout this paper $W(r)$ will denote a standard univariate Brownian Motion and $\widetilde{W}^0(.)$ a p dimensional Brownian Bridge.

ASSUMPTIONS: *(A1)* e_t *is an* $i.i.d(0, \sigma_e^2)$ *random variable satisfying the Functional Central Limit Theorem* $\sum_{t=1}^{[Tr]} e_t/\sqrt{T} \Rightarrow \sigma_e W(r)$, *(A2)* y_t *is such that* $\Delta y_t = \alpha + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + e_t$ *with* $\Gamma(z) = 1 - \gamma_1 z - \ldots - \gamma_p z^p$ *having all its roots lie outside the complex unit circle.*

With the above assumptions we are now in a position to state our main result about the limiting distributions of $SupWald^A$ and $SupWald^B$. Note that the underlying DGP may have a nonzero drift since our fitted specification contains a deterministic time trend but it can obviously not contain any deterministic trend components. To economise on notation we let $\overline{W}(r) = (1\ r\ W(r))'$, $M(\pi) = \int_0^\pi \overline{W}(r)\overline{W}(r)'dr$ and $M(\pi)^* = M(\pi) - M(\pi)M(1)^{-1}M(\pi)$. We also let ADF_∞ refer to the standard unit root limiting distribution of the t-ratio under the scenario of a random walk with drift in the DGP and a fitted model with a constant and trend (see Hamilton (1994, pp. 497-500)).

PROPOSITION 1: *Under (A1) and (A2) and as* $T \rightarrow \infty$ *we have*

$$\sup_{\pi \in [\pi_1, \pi_2]} W_T^A(\pi) \Rightarrow ADF_\infty^2 + \sup_{\pi \in [\pi_1, \pi_2]} [Q_1(\pi) + Q_2(\pi)] \tag{4}$$

$$\sup_{\pi \in [\pi_1, \pi_2]} W_T^B(\pi) \Rightarrow \sup_{\pi \in [\pi_1, \pi_2]} [Q_1(\pi) + Q_2(\pi)] \tag{5}$$

where

$$Q_1(\pi) = \left[\int_0^\pi \overline{W}dW - M(\pi)M(1)^{-1}\int_0^1 \overline{W}dW\right]' M^*(\pi)^{-1}\left[\int_0^\pi \overline{W}dW - M(\pi)M(1)^{-1}\int_0^1 \overline{W}dW\right]$$

and

$$Q_2(\pi) = \frac{\widetilde{W}^0(\pi)'\widetilde{W}^0(\pi)}{\pi(1-\pi)}.$$

It is important to first note that both distributions in (4) and (5) are free of any nuisance parameters and can easily be tabulated across alternative magnitudes of p and possible choices of π. If the model does not include any lagged dependent regressors (4) and (5) continue to hold as stated but without the $Q_2(\pi)$ component.

The limiting random variable in (5) has two components with the first one arising due to the presence of a nonstationary regressor and deterministic trend components while the second one given by the normalised quadratic form in Brownian Bridges is induced by the inclusion of the p stationary regressors z_{t-1} in the right hand side of (1). This latter component $Q_2(\pi)$ is well known in the literature on testing for structural breaks within purely stationary environments (see Andrews (1992), Hansen (1997) amongst others) while the first component $Q_1(\pi)$ is novel and nonstandard, arising due to the joint interaction of breaks and unit roots. When the model in (1) contains no lagged dependent regressors (i.e. when $p = 0$) we have $\sup_\pi W_T^A(\pi) \Rightarrow \sup_\pi Q_1(\pi)$.

At this stage it is interesting to highlight the fact that controlling for the number of parameters whose stability is being tested, $\sup_\pi Q_1(\pi)$ lies far off to the right of its counterpart arising under pure stationarity (i.e $\sup_\pi Q_2(\pi)$). This can be observed through a comparison of the simulated quantiles of (5) presented in Table 2 below under $p = 0$ with those in Table 1 of Andrews (1992). An immediate implication of this observation is the fact that ignoring a unit root when conducting inferences about structural breaks under a wrongly assumed stationary setting will systematically lead to the detection of spurious breaks.

It is also interesting to contrast the formulation in (5) with its counterpart occurring when the regimes are determined by a stationary threshold variable as in Caner and Hansen (2001) instead of time itself. Ignoring the presence of lagged dependent regressors for instance we have $\sup_\pi W_T^A(\pi) \Rightarrow \sup_\pi Q_1(\pi)$ which can be contrasted with a limit of the form $\sup_\pi Q_2(\pi)$ that arises in the threshold setting for the same set of regressors (see Proposition 3 in Pitarakis (2008)). Under both scenarios the limits are clearly free of nuisance parameters as long as the fitted model does not contain stationary regressors. When such regressors are included our limit in (5) continues to be free of nuisance parameters while in the threshold setting the limiting random variable becomes a complicated function of unknown model specific moments.

4 Tabulations and Experimental Illustrations

4.1 Empirical Quantiles

Our initial objective is to provide a tabulation of the limiting distributions presented in (4) and (5). We take $\Delta y_t = \alpha + e_t$ as our DGP and with no loss of generality set $\alpha = 0$. Note that both distributions depend on p the number of parameters associated with the stationary lagged dependent regressors whose stability is being tested and $\pi_1 = 1 - \pi_2$ which we set at 10% following standard practice. Results across key quantiles and magnitudes of p are presented in Table 1 below for $SupWald^A$ and in Table 2 for $SupWald^B$. All our experiments are conducted using $N = 5000$ replications and take $e_t \equiv NID(0,1)$ throughout.

Table 1. Quantiles of $SupWald^A$

	$T = 200$			$T = 400$			$T = 1000$		
	90%	95%	97.5%	90%	95%	97.5%	90%	95%	97.5%
$p = 0$	29.66	32.24	34.47	30.23	32.59	35.09	30.80	33.28	35.57
$p = 1$	30.99	33.81	36.11	31.52	33.92	36.52	32.22	34.74	37.32
$p = 2$	32.45	35.24	37.94	32.96	35.67	37.94	33.53	36.09	38.76
$p = 3$	33.89	36.52	39.28	34.49	37.34	39.55	35.01	37.69	40.22
$p = 4$	35.04	38.06	40.68	35.66	38.15	40.52	36.43	38.92	41.61
$p = 5$	36.97	39.81	42.16	37.08	39.68	42.74	37.67	40.43	42.88
$p = 6$	37.82	40.67	43.63	38.65	41.44	43.98	39.07	41.84	44.40
$p = 7$	39.00	41.93	45.38	39.91	42.68	45.17	40.35	43.41	46.11
$p = 8$	40.21	43.39	46.24	40.73	43.69	46.85	41.67	44.36	47.11

Looking first at the variation in critical values across different sample sizes we note that the T=200 and T=400 based finite sample distributions lie slightly to the left of their asymptotic counterpart as proxied by T=1000. Although the relevant quantiles remain numerically very close and thus distortions should remain limited when basing finite sample inferences on asymptotic quantiles the above figures suggest that the test based on $SupWald^A$ may be slightly undersized in small samples.

We also repeated the above exercise for the finite sample distributions of $SupWald^B$ whose quantiles are presented in Table 2 below. The simulated quantiles continue to suggest that inferences based on moderately sized samples should be sufficiently accurate even when the hypotheses being tested involve a large number of parameters but our earlier discussion about the potential finite sample undersizeness remains valid for $SupWald^B$ as well. The row labelled $p = 0$ corresponds to a scenario where the ADF regression contains no lagged dependent regressors so that the relevant limiting distribution is in fact given by $\sup_\pi Q_1(\pi)$ in (5). It is interesting to note that this limiting distribution lies markedly to the right of that of $\sup_\pi Q_2(\pi)$ which is commonly used when testing for structural breaks in stationary settings and whose quantiles across different magnitudes of p are available from Andrews (1992). This is an important and useful observation since it points to a spurious detection of a break phenomenon when a model contains a unit root variable but its presence is ignored and inferences are conducted using the Brownian Bridge asymptotics that are valid solely under stationarity.

Table 2. Quantiles of $SupWald^B$

	$T = 200$			$T = 400$			$T = 1000$		
	90%	95%	97.5%	90%	95%	97.5%	90%	95%	97.5%
$p = 0$	24.64	27.34	29.81	25.14	27.73	30.23	25.97	28.71	31.27
$p = 1$	26.21	28.88	31.54	26.63	29.31	31.88	27.25	29.88	32.48
$p = 2$	27.72	30.66	33.31	27.97	30.72	33.25	28.56	31.17	33.72
$p = 3$	29.27	32.23	35.09	29.56	32.36	35.10	29.99	32.84	35.38
$p = 4$	30.21	33.25	36.18	30.86	33.90	36.78	31.29	34.17	36.69
$p = 5$	31.84	34.89	37.58	32.00	34.94	37.55	32.70	35.51	38.06
$p = 6$	32.90	35.99	38.56	33.32	36.08	38.79	34.21	37.00	39.31
$p = 7$	34.41	37.26	40.75	34.90	37.90	40.68	35.37	38.09	41.18
$p = 8$	35.41	38.43	41.58	35.98	39.27	41.96	36.61	39.48	41.97

The above observations suggest that under small to moderate sample sizes it may be preferable to use our finite sample quantiles obtained under $T = 200$ or $T = 400$. To gain further insight into the size properties of our two tests when using the $T = 1000$ based quantiles, Table 3 below presents various empirical size estimates across different magnitudes of p for 2.5% and 5% nominal levels. The critical values are those displayed under $T = 1000$ in Tables 1-2 above.

Table 3. Empirical Size Properties of $SupWald^A$ and $SupWald^B$

	$SupWald^A$						$SupWald^B$					
	$T = 200$		$T = 400$		$T = 600$		$T = 200$		$T = 400$		$T = 600$	
	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%	2.5%	5.0%
$p = 0$	1.82	3.72	2.12	4.16	2.62	4.66	1.68	3.36	1.92	4.10	2.00	4.50
$p = 1$	2.22	4.60	2.42	4.74	2.58	5.08	2.18	4.34	2.48	4.74	2.36	4.92
$p = 2$	2.26	4.84	2.50	4.90	2.54	5.66	2.22	4.58	2.62	5.06	2.60	5.48

The T=200 based figures displayed in Table 3 above confirm our earlier discussion about the undersizeness of the two tests when inferences are based on asymptotic quantiles. As T is allowed to grow however and taking simulation variation into account we note that empirical sizes match closely their nominal counterparts for both test statistics.

4.2 Power Properties

Here we explore the power properties of our test statistics across a range of fixed departures from the null hypotheses of interest. Our experiments are conducted across samples of size $T = 200$ and $T = 400$ and make use of the corresponding quantiles displayed in Tables 1-2. Our most general DGP is given by $\Delta y_t = (\alpha_1 + \rho_1 y_{t-1} + \gamma_{11}\Delta y_{t-1})I_{1t} + (\alpha_2 + \rho_2 y_{t-1} + \gamma_{21}\Delta y_{t-1})I_{1t} + e_t$ so that we concentrate on a $p = 1$ scenario and the following parameterisations

Table 4. DGP Parameterisations

	α_1	α_2	ρ_1	ρ_2	γ_{11}	γ_{21}
M_1	0.0	0.0	0.0	-0.1	0.5	0.5
M_2	0.0	0.0	0.0	-0.2	0.5	0.5
M_3	0.0	0.0	-0.1	0.0	0.5	0.5
M_4	0.0	0.0	-0.2	0.0	0.5	0.5
M_5	0.0	0.0	-0.1	0.0	-0.1	0.3
M_6	0.0	0.0	-0.1	0.0	-0.1	0.5
M_7	0.0	0.0	0.0	0.0	-0.1	0.3
M_8	0.0	0.0	0.0	0.0	-0.1	0.5
M_9	0.0	0.0	-0.1	-0.1	-0.1	0.3
M_{10}	0.0	0.0	-0.2	-0.2	-0.1	0.5.

The above models cover a wide range of scenarios with a particular focus on breaks in the $\rho's$ and $\gamma's$. We are particularly interested in assessing the ability of a test statistic such as $SupWald^A$ to detect switches in the $\rho's$ from unit root to stationarity and vice-versa (models M_1 to M_6). All our specifications have the structural break occur at $\pi_0 = 0.5$. Models M_7 and M_8 are AR models in first differences having their slope parameters shift following a structural break while models M_9 and M_{10} are AR(2) specifications that are stationary within each regime.

Table 5 below presents the correct decision frequencies corresponding to our two test statistics. Its last column also includes the corresponding outcomes for the squared version of the standard ADF t-statistic of the unit root hypothesis. Note that our power figures have been computed using the correct critical values obtained under $T = 200$ and $T = 400$. In the case of the ADF based squared t-ratio we also obtained its quantiles through simulations. More specifically, its 90%, 95% and 97.5% quantiles are given by $\{9.85, 11.73, 13.73\}$, $\{9.96, 11.77, 13.62\}$ and $\{9.89, 11.76, 13.87\}$ for $T = 200, 400$ and $T = 1000$

respectively. At this stage it is important to reiterate that it is not our aim to view our tests as alternatives to the standard ADF test and Table 5 below includes its performance solely for the purpose of gauging the usefulness of our Sup based tests under particular scenarios.

Table 5. Empirical Power

	$SupWald^A$		$SupWald^B$		$Wald\ ADF(\rho=0)$	
	$T=200$	$T=400$	$T=200$	$T=400$	$T=200$	$T=400$
M_1	39.00	80.08	29.12	59.84	12.70	18.48
M_2	81.28	99.86	61.81	96.12	74.00	23.80
M_3	10.78	44.12	6.84	28.26	11.24	15.54
M_4	43.64	97.84	29.72	89.70	14.68	17.12
M_5	26.28	78.92	26.28	74.24	2.36	3.48
M_6	70.68	99.84	71.52	99.32	1.02	1.26
M_7	16.50	45.08	17.46	45.44	1.26	1.26
M_8	54.82	94.86	57.32	94.86	0.64	0.54
M_9	36.08	94.38	6.94	25.16	47.00	99.12
M_{10}	98.00	100.00	39.90	92.16	99.56	100.00

It is interesting to note that our $SupWald^A$ statistic is able to successfully detect departures from the null when the shift affects solely ρ_1 or ρ_2. Under M_2 for instance $SupWald^A$ is able to correctly reject the null close to 100% of the times under $T = 400$ and is characterised by equally powerful outcomes across most other configurations. It is also important to note a marked difference in behaviour when the $\rho_i's$ switching from a unit root type of behaviour to stationarity (e.g. M_1 and M_2) as opposed the $\rho_i's$ switching from stationarity to a unit root region (e.g. M_3, M_4, M_5 and M_6). Across smaller sample sizes our $SupWald^A$ based test is substantially more powerful in detecting departures from the null such as M_1 or M_2 than when the model switches from stationarity to a unit root scenario. This phenomenon can clearly be observed when comparing M_1 with M_3. Overall and looking across all our scenarios we note a good to excellent ability of $SupWald^A$ to detect departures from the null along a variety of directions.

At this stage it is also interesting to compare our joint tests with the behaviour of the standard ADF statistic. As expected we first note that under models such as M_9 and M_{10} which have both their autoregressive roots outside the unit circle, the ADF statistic displays good power properties similar in magnitude to the correct decision frequencies characterising $SupWald^A$. For most other scenarios we note that ADF based inferences are mostly unable to move away from the unit root null even with moderately large samples. This is perhaps not surprising since fundamentally it is not designed to handle cases such as $M_1 - M_6$ and if one is interested in exploring the presence of such scenarios in the data then our proposed test statistic appears to be particularly suitable.

Cases M_7 and M_8 are also interesting. They correspond to DGPs with a unit root throughout but shifts in the parameters corresponding to the stationary regressors. In this instance the ADF is rightly unable to move away from the unit root null while our $SupWald^A$ based inferences also rightly lead to rejections of the joint null under T=400 in particular due to the presence of a break.

5 Conclusions and Further Remarks

We proposed test statistics designed to test the joint hypothesis of a unit root *and* parameter stability in the context of an autoregressive model. Their limiting distributions were shown to be free of nuisance parameters and easily tabulated. Finally through a set of numerical experiments we illustrated their usefulness for detecting a wide range of departures from the null hypotheses of interest. Although our probabilitic framework is sufficiently general to allow our proposed toolkit wide applicability numerous extensions such as the inclusion of further breaks, possible regime dependent heteroskedasticity etc. are also possible and would be interesting to pursue. It is also important to emphasise that our use of a series of test statistics such as $SupWald^A$, $SupWald^B$ and the standard ADF based unit root test should not be seen as an attempt to include them within a sequential testing strategy. Each test has its own merit and may be considered individually depending on the application in hand. Combining inferences from different tests that may or may not be correlated is a notoriously difficult problem which is beyond the scope of this paper. Even under independence which would allow one to control the overall size of a sequentially implemented test the choice of individual significance levels is not obvious and may lead to very different conclusions.

APPENDIX

PROOF OF PROPOSITION 1. We focus on the proof of (4) since that of (5) is included in our results below. Recall that our maintained model is as described in Assumption (A2) and given by $\Delta y_t = \alpha + \sum_{j=1}^{p} \gamma_j \Delta y_{t-j} + e_t$ which for greater convenience we also reformulate as $\Delta y_t = \mu + u_t$ with $u_t = \sum_{j=1}^{p} \gamma_j u_{t-j} + e_t$ and $\mu = \alpha/(1 - \gamma_1 - \ldots - \gamma_p)$. Our fitted model is given by (1). Setting $\zeta_t = y_t - \mu t = \sum_{j=1}^{t} u_j$ we rewrite it as

$$\Delta y_t = \begin{cases} \alpha_1 + \beta_1 t + \rho_1(y_{t-1} - \mu(t-1)) + \sum_{j=1}^{p} \gamma_{1j}(\Delta y_{t-j} - \mu) + \rho_1 \mu(t-1) + \mu \sum_{j=1}^{p} \gamma_{1j} + e_t & t \leq k \\ \alpha_2 + \beta_2 t + \rho_2(y_{t-1} - \mu(t-1)) + \sum_{j=1}^{p} \gamma_{2j}(\Delta y_{t-j} - \mu) + \rho_2 \mu(t-1) + \mu \sum_{j=1}^{p} \gamma_{2j} + e_t & t > k \end{cases} \tag{6}$$

and more compactly as

$$\Delta y_t = \begin{cases} \alpha_1^* + \beta_1^* t + \rho_1 \zeta_{t-1} + \sum_{j=1}^{p} \gamma_{1j} u_{t-j} + e_t & t \leq k \\ \alpha_2^* + \beta_2^* t + \rho_2 \zeta_{t-1} + \sum_{j=1}^{p} \gamma_{2j} u_{t-j} + e_t & t > k. \end{cases} \tag{7}$$

Letting $X^* = (1\ t\ \zeta_{t-1}\ u_{t-1} \ldots u_{t-p})$ and $X_i^* = (1\ t\ \zeta_{t-1}\ u_{t-1} \ldots u_{t-p}) I_i$ for $i = 1, 2$ we can write (7) as $\Delta Y = X_1^* \theta_1^* + X_2^* \theta_2^* + e = X^* \theta_2^* + X_1^*(\theta_1^* - \theta_2^*) + e$. Here $\theta_i^* = (\alpha_i^*, \beta_i^*, \rho_i, \gamma_{i1}, \ldots, \gamma_{ip})$. Letting $\Psi^* = \theta_1^* - \theta_2^*$ it is now immediately apparent that testing H_0^A in (1) is equivalent to testing $H_0^{A*} : \Psi^* = 0, \rho_1 = 0$ in our reparameterised model. With W stacking the elements of $(1\ t\ u_{t-1} \ldots u_{t-p})$ and defining $M_W = I - W(W'W)^{-1}W'$ as well as $M_{X^*, X_1^*} = M_{X^*} - M_{X^*} X_1^* (X_1^{*\prime} M_{X^*} X_1^*)^{-1} X_1^* M_{X^*}$ it follows that the Wald statistic can be written as $W_T^A(k) = [\Delta Y' M_W \Delta Y - \Delta Y' M_{X^*, X_1^*} \Delta Y]/\hat{\sigma}_e^2$ and imposing the null hypothesis leads to $W_T^A(k) = [e' M_W e - e' M_{X^*, X_1^*} e]/\hat{\sigma}_e^2$. Here $\hat{\sigma}_e^2 = \Delta Y' \Delta Y - \sum_{i=1}^{2} \Delta Y' X_i^* (X_i^{*\prime} X_i^*)^{-1} X_i^{*\prime} \Delta Y$. Before proceeding with the limiting behaviour of $W_T^A(k)$ it is convenient to reformulate our test statistic as

$$W_T^A(k) \quad = \quad \frac{1}{\hat{\sigma}_*^2} \left(e' M_W e - e' M_{X^*} e \right) \frac{\hat{\sigma}_*^2}{\hat{\sigma}_e^2} + \frac{1}{\hat{\sigma}_e^2} \left(e' M_{X^*} e - e' M_{X^*, X_1^*} e \right). \tag{8}$$

Within the above formulation it is easy to note that the first component in the right hand side of (8) does not depend on k and corresponds to a Wald statistic for testing the null of a unit root within an ADF specification that includes a constant and trend and when the underlying model is a random walk with drift. From the WLLN it is also clear that $\hat{\sigma}_*^2/\hat{\sigma}_e^2 \xrightarrow{p} 1$ hence establishing the ADF_∞^2 limit for the first component in the right hand side of (8) (see Hamilton (1994), pp. 497-500 for a more explicit formulation of ADF_∞). The use of the Continuous mapping theorem combined with the following intermediate distributional results will then lead to our representation in (4). For simplicity and no loss of generality we set $\sigma_e^2 = 1$ throughout and let $D_T = diag(\sqrt{T}, T^{3/2}, T, \sqrt{T}, \ldots, \sqrt{T})$ refer to a suitable $(p+3) \times (p+3)$ diagonal normalisation matrix. We first note that $e' M_{X^*} e - e' M_{X^*, X_1^*} e = e' M_{X^*} X_1^* (X_1^{*\prime} M_{X^*} X_1^*)^{-1} X_1^{*\prime} M_{X^*} e =$

$[e'X_1^* - e'X^*(X^{*\prime}X^*)^{-1}X_1^{*\prime}X_1^*][X_1^{*\prime}X_1^* - X_1^{*\prime}X_1^*(X^{*\prime}X^*)^{-1}X_1^{*\prime}X_1^*]^{-1}[X_1^{*\prime}e - X_1^{*\prime}X_1^*(X^{*\prime}X^*)^{-1}X^{*\prime}e]$. Using standard unit root asymptotics (see Hamilton (1994, pp.)) the CMT together with the fact that with u stacking the elements of $(u_{t-1}, \dots, u_{t-p})$ under our assumptions the Ergodic Theorem ensures that $u'u/T \stackrel{p}{\to} E[u'u] \equiv V > 0$ we have

$$D_T^{-1}[X_1^{*\prime}M_{X^*}X_1^*]D_T^{-1} \quad \Rightarrow \quad \begin{pmatrix} M^*(\pi) & 0 \\ 0' & \pi(1-\pi)V \end{pmatrix} \tag{9}$$

and

$$D_T^{-1}(X_1^{*\prime}X_1^*)(X^{*\prime}X^*)^{-1}D_T^{-1} \quad \Rightarrow \quad \begin{pmatrix} \int_0^\pi \overline{W}\overline{W}' \left(\int_0^1 \overline{W}\overline{W}'\right)^{-1} & 0 \\ 0' & \pi I_p \end{pmatrix}. \tag{10}$$

Next we note that

$$\mathrm{D}_T^{-1}X_1^{*\prime}e = (\textstyle\sum_{t=1}^{T\pi} e_t/\sqrt{T}, \sum_{t=1}^{[T\pi]} te_t/T^{3/2}, \sum_{t=1}^{[T\pi]} \zeta_{t-1}e_t/T, \sum_{t=1}^{[T\pi]} u_{t-1}e_t/\sqrt{T}, \dots, \sum_{t=1}^{[T\pi]} u_{t-p}e_t/\sqrt{T})'.$$

so that standard CLT and FCLT based arguments lead to

$$D_T^{-1}X_1^{*\prime}e \quad \Rightarrow \quad \begin{pmatrix} \int_0^\pi \overline{W}dW(r) \\ V^{1/2}\widetilde{W}_p(\pi) \end{pmatrix}. \tag{11}$$

with $\widetilde{W}_p(\pi)$ denoting a p-dimensional standard Brownian Motion. Similarly, we have

$$D_T^{-1}X^{*\prime}e \quad \Rightarrow \quad \begin{pmatrix} \int_0^1 \overline{W}dW(r) \\ V^{1/2}\widetilde{W}_p(1) \end{pmatrix} \tag{12}$$

so that combining (9)-(11) leads to

$$X_1^{*\prime}e - X_1^{*\prime}X_1^*(X^{*\prime}X^*)^{-1}X^{*\prime}e \quad \Rightarrow \quad \begin{pmatrix} \int_0^\pi \overline{W}dW(r) - M(\pi)M(1)^{-1}\int_0^1 \overline{W}dW(r) \\ V^{1/2}[\widetilde{W}_p(\pi) - \pi\widetilde{W}_p(1)] \end{pmatrix}. \tag{13}$$

Combining (9) and (13) then leads to the desired result.

REFERENCES

Andrews, D. W. K., (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, Vol. 61, pp. 821-856.

Bai, J. (1998), "A Note On Spurious Break," *Econometric Theory*, Vol. 14, pp. 663-669.

Banerjee, A., R. L. Lumsdaine and Stock, J. H. (1992), "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," *Journal of Business Economic Statistics*, Vol. 10, pp 271-87.

Caner, M. and Hansen, B. E. (2001), "Threshold Autoregression with a Unit Root," *Econometrica,* Vol. 69, pp. 1555-1596.

Hansen, B. E. (1997), "Approximate asymptotic p-values for structural change tests," *Journal of Business and Economic Statistics*, Vol. 15, pp. 60-67.

Hamilton, J. D. (1994), Time Series Analysis, Princeton University Press.

Harris, D., Harvey, D. I., Leybourne, S. J. and Taylor, A. M. R. (2009), "Testing For A Unit Root In The Presence Of A Possible Break In Trend," *Econometric Theory*, Vol. 25, pp. 1545-1588.

Kim, D. and Perron, P. (2009), "Unit root tests allowing for a break in the trend function at an unknown time under both the null and alternative hypotheses," *Journal of Econometrics*, Vol. 148, pp. 1-13.

Perron, P. (1989), "The Great Crash, the Oil Price Shock, and the Unit Root Hypothesis," *Econometrica,* Vol. 57, pp. 1361-1401.

Pitarakis, J. (2008), "Threshold Autoregressions with a unit root: Comment," *Econometrica*, Vol. 76, pp. 1207-1217.

Zivot, E. and Andrews, D. W. K (1992), "Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis," *Journal of Business Economic Statistics, Vol. 10, pp. 251-70.*