Valid Inference in Partially Unstable Generalized Method of Moments Models

HONG LI

Brandeis University

and

ULRICH K. MÜLLER

Princeton University

This paper considers time series Generalized Method of Moments (GMM) models where a subset of the parameters are time varying. We focus on an empirically relevant case with moderately large instabilities, which are well approximated by a local asymptotic embedding that does not allow the instability to be detected with certainty, even in the limit. We show that for many forms of the instability and a large class of GMM models, usual GMM inference on the subset of stable parameters is asymptotically unaffected by the partial instability. In the empirical analysis of presumably stable parameters—such as structural parameters in Euler conditions—one can thus ignore moderate instabilities in other parts of the model and still obtain approximately correct inference.

1. INTRODUCTION

Instabilities in the parameters of econometric time series models are a plausible and empirically widespread phenomenon. Time-varying market conditions, rules and regulations, and technological innovations change the economic environment. As pointed out by Lucas (1976), these environmental changes induce behavioural changes of rational economic agents, which result in time-varying parameters in many econometric relationships. In addition, misspecifications of econometric models can also manifest themselves in the form of time-varying parameters. Empirically, Ghysels (1998), Stock and Watson (1996), Boivin (1999), and Cogley and Sargent (2005), for instance, find instabilities in macro-economic and finance relationships.

Econometric theory has focussed to a large extent on the problem of testing the null hypothesis that a time series model is stable over time against the alternative of parameter variation whose exact form is unknown: see, for instance, Nyblom (1989), Andrews (1993), Andrews and Ploberger (1994), Sowell (1996), Bai and Perron (1998), Hansen (2000), Andrews (2003), and Elliott and Müller (2006) for some recent contributions. Much less work is concerned with the next step: what is one to do once instabilities are suspected? One useful result, established in Bai (1994) and generalized in Bai and Perron (1998), concerns inference in linear regressions with a discrete number of parameter shifts at unknown times. If the parameter shifts are large in the sense that reasonable tests detect the instability with probability 1 in the limit, then standard inference on the coefficients in the various regimes remains asymptotically valid when the regime dates are based on least-squares break date estimators.

In many applications, however, instabilities are arguably not sufficiently large for these asymptotics to generate reliable approximations: Linde (2001) for instance argues that economically important changes in monetary policy lead to parameter instabilities that are small in the sense of being difficult to detect empirically. More generally, the instabilities in bivariate relationships between macro-economic data series documented in Stock and Watson (1996) are often only borderline significant. In such instances, more accurate approximations are generated by asymptotics where reasonable tests detect the instabilities with probability strictly smaller than 1 in the limit (see Elliott and Müller, 2007, for some quantitative results on break date confidence sets).

In this paper, we analyse models where only a subset of parameters is unstable and focus on such small instabilities. We ask the question: how to conduct valid inference on the stable subset of parameters. The answer turns out to be more straightforward than it might seem. For a very wide range of unstable parameter paths, and for a large class of Hansen's (1982) Generalized Method of Moments (GMM) models, standard GMM inference (ignoring the partial instability) remains asymptotically valid for the subset of stable parameters.

To develop some intuition for this result, consider the linear model:

$$Y_t = X_t\theta_{1,t} + Z_t\theta_2 + \varepsilon_t, \ \varepsilon_t \sim \text{i.i.d.}(0,\sigma^2), \tag{1}$$

where X_t and Z_t are two (possibly correlated) scalar random variables. By standard OLS algebra (the Frisch–Waugh Theorem), the t statistic on θ_2 in a regression of Y_t on (X_t, Z_t) is numerically identical to the t statistic in a regression of Y_t on $(X_t, \tilde{Z}_t)$, where $\tilde{Z}_t = Z_t - \tilde{\beta}X_t$ is the residual of a regression of Z_t on X_t. In terms of $\tilde{Z}_t$, equation (1) becomes:

$$Y_t = X_t(\theta_{1,t} + \tilde{\beta}) + \tilde{Z}_t\theta_2 + \varepsilon_t = X_t(\theta_{1,0} + \tilde{\beta}) + \tilde{Z}_t\theta_2 + X_t(\theta_{1,t} - \theta_{1,0}) + \varepsilon_t. \tag{2}$$

If (X_t, Z_t) is approximately stationary in the sense that $T^{-1}\sum_{t=1}^T X_t\tilde{Z}_t = 0$ also implies $T^{-1}\sum_{t=1}^{[sT]} X_t\tilde{Z}_t \xrightarrow{p} 0$ for all $s \in [0,1]$, and $\theta_{1,t} - \theta_{1,0}$ is a smooth function of t, then the additional "error term" $X_t(\theta_{1,t} - \theta_{1,0})$ is approximately orthogonal to $\tilde{Z}_t$. If in addition, the amount of parameter instability is small in the sense that $\sup_{t\leq T}\|\theta_{1,t} - \theta_{1,0}\| = O(T^{-1/2})$, then the error term ε_t dominates the randomness of the OLS estimator of θ_2, and inference on θ_2 remains largely unaffected by the instability of θ_1. In contrast, if (X_t, Z_t) is a persistent series, lack of correlation between $\tilde{Z}_t$ and X_t does not imply lack of correlation between $X_t(\theta_{1,t} - \theta_{1,0})$ and $\tilde{Z}_t$, and the presence of $X_t(\theta_{1,t} - \theta_{1,0})$ invalidates standard inference for θ_2, even for small parameter instabilities.

The same result also holds for more general GMM inference, using the usual linearization arguments. The appropriate analogy for the constraint that (X_t, Z_t) are not too persistent is the assumption that sample averages of the derivative of the moment condition are approximately the same in all parts of the sample. This holds for most linear or non-linear globally stationary models, such as stationary vector autoregressive models. It typically fails to hold, though, for models that generate deterministically or stochastically trending data.

A leading economic example of a partially unstable GMM model is Euler moment conditions of optimizing agents under a time-varying policy environment. Rational economic agents adapt their optimal behaviour to policy changes. Econometrically, this leads to reduced-form equations that exhibit time-varying parameters. At the same time, structural parameters describing preferences and technology might very well remain constant, and their values are crucial for conducting proper policy analysis. For concreteness, consider a stylized two-equation system involving (i) a New Keynesian Phillips Curve (NKPC), which is a rational expectations Euler condition in inflation and unemployment gap, and (ii) a reduced-form process for the unemployment gap, the driving variable of the NKPC (see Blanchard and Gali, 2007, for the theoretical

derivation of this specification). Expressing the NKPC in first differences and with an AR(2) specification for the unemployment gap, we obtain:

$$\Delta \pi_t = \phi E_t \Delta \pi_{t+1} + \kappa s_t + \varepsilon_t, \tag{3}$$

$$s_t = \rho_{1,t} s_{t-1} + \rho_{2,t} s_{t-2} + \xi_t, \tag{4}$$

where π_t and s_t are the inflation rate and unemployment gap, respectively; E_t is the conditional expectation at date t; and the disturbance terms ε_t and ξ_t are i.i.d. mean zero. In this example, economic theory has direct implications for the stability of the various parameters: the coefficients ρ_1 and ρ_2 are functions of current monetary policy. With a time-varying monetary policy, ρ_1 and ρ_2 therefore become unstable. The Euler equation (3), in contrast, is derived from the economic agents' optimization problem. As long as preferences and technology remain constant through time, economic theory implies ϕ and κ to be stable, even in the face of a time-varying monetary policy. Our results imply that one might estimate the systems (3) and (4) by standard GMM, ignoring the time variability in ρ_1 and ρ_2, and one still obtains approximately valid inference about the structural parameters ϕ and κ, at least as long as the time variability in ρ_1 and ρ_2 is not too pronounced. We return to this example in Section 4 and find that this is a reasonable approximation in a Monte Carlo exercise calibrated to U.S. data. Also see Li (2008) for an empirical application of our results to an investment model.

We also find that popular tests of stability of a subset of parameters are typically affected by instabilities in the non-tested parameters. An additional contribution of this paper is the derivation of a class of modified tests whose rejection probability is unaffected by local instabilities in the non-tested parameters.

Our results allow for parameter instabilities of a magnitude that corresponds to local alternatives of efficient stability tests. Formally, in such asymptotics, the magnitude of the instability is of the order $T^{-1/2}$ in a sample of size T. As noted above, this does not mean that our results only apply to economically insignificant instabilities. In many applications, there is substantial uncertainty about the presence and form of instabilities. Accurate approximations are then generated by a modelling strategy in which there is only limited information about the instability asymptotically, as in the $T^{-1/2}$ neighbourhood.

What is more, from a more theoretical perspective, it makes sense to focus on local deviations from standard model assumptions in a robustness analysis. After all, when parameter instabilities are large, the problem can be detected consistently with an appropriate test and, at least for a finite number of discrete shifts, the inclusion of the appropriate dummies leads to valid inference, as demonstrated by Bai and Perron (1998). In contrast, when parameter instabilities are of the order $T^{-1/2}$, there is no way of knowing for sure whether the parameters are unstable, and there is no obvious remedy if one believes they are. Our results precisely cover this latter case, where it is challenging to derive more immediate approaches to time-varying nuisance parameters.

In applications of our results, one must decide whether these asymptotic considerations yield accurate approximations. This is, of course, true of all asymptotically justified inference. But the problem is more severe here since it is not hard to see that in general, our results do not hold for asymptotics with a fixed magnitude of the time variation,[1] where, for instance,

---

1. As noted above, inference based on the results of Bai and Perron (1998) face the same problem in reverse—their approach yields poor approximations for local parameter instabilities. See Elliott and Müller (2007) for details.

$\theta_{1,t} - \theta_{1,0} = \mathbf{1}[t \geq T/2]$ rather than $\theta_{1,t} - \theta_{1,0} = T^{-1/2}\mathbf{1}[t \geq T/2]$ in equation (1).[2] A formal solution is to reject the null hypothesis for the value of a parameter of interest using the usual GMM Wald test only if the observed p value of a good parameter stability test is larger than some threshold $\varepsilon > 0$ and to never reject otherwise. Since the p value of a good parameter stability test converges to 0 in probability for instabilities that are larger than the $T^{-1/2}$ neighbourhood, this approach yields uniformly valid inference over all magnitudes of the partial instability (at the cost of being potentially very conservative for large instabilities). But in practice, picking ε too small relative to the actual sample size might still yield poor approximations, so that a judgement must still be made. In Section 3, we provide small sample results for two data generating processes to shed some light on this issue, and we find that the local asymptotics considered in this paper tend to generate accurate approximations for sample sizes and amounts of instability that are plausibly encountered in macro-economic and financial applications.

On a technical level, the analysis of time series models with time-varying parameters faces the difficulty that these models tend to generate non-stationary data. This complicates the justification of asymptotic approximations, such as those generated from laws of large numbers. We address these difficulties by providing sufficient conditions for the unstable model to be *contiguous* to the corresponding stable model. In the analysis of parameter stability tests for fully specified parametric models, the concept of contiguity has been employed before in Andrews and Ploberger (1994) and Elliott and Müller (2006), although these papers address more specific forms of parameter instability than considered here. Contiguity ensures that approximation errors that are $o_p(1)$ in the stable model remain $o_p(1)$ in the corresponding unstable model. It therefore suffices to make appropriate assumptions on the stable model, and derive the corresponding properties of the unstable model via contiguity. The results we establish with this indirect reasoning might be of independent interest; they might be used, for instance, to justify the efficient tests of parameter instability in Sowell (1996), the median unbiased estimators of the amount of local instability in Stock and Watson (1998) or the inference for the date of a single break in linear regression in Elliott and Müller (2007) for a wide class of data generating processes, and results of this paper have been used in the derivation of efficient parameter path estimators in Müller and Petalas (2007).

The next section contains the asymptotic results of this paper. Section 2.1 introduces the model and discusses a high-level condition on the partially unstable GMM model. The high-level condition is sufficient for the main asymptotic result presented in Section 2.2, followed by a discussion of its implications for econometric practice. In Section 2.4, we derive contiguity-based arguments to justify the high-level condition on the unstable model by appropriate assumptions about the properties of the corresponding stable model. Section 3 contains a Monte Carlo study that investigates the small sample relevance of the asymptotic results. Section 4 concludes. Proofs are collected in the Appendix.

2. ASYMPTOTIC RESULTS

2.1. *Model and high-level condition*

Consider a GMM model with the unknown $m \times 1$ parameter vector θ, an element of the parameter space $\Theta \subset \mathbb{R}^m$. The observed data in a sample of size T are given by a triangular array of

---

2. For such a fixed parameter instability $\theta_{1,t} - \theta_{1,0} = \mathbf{1}[t \geq T/2]$, the outcome depends strongly on properties of $\{(X_t, Z_t)\}_{t=1}^T$: For $\{(X_t, Z_t)\}_{t=1}^T$ i.i.d. and independent of $\{\varepsilon_t\}_{t=1}^T$, the heteroskedasticity robust t statistic on θ_2 remains *unconditionally* asymptotically valid but not conditionally on $\{(X_t, Z_t)\}_{t=1}^T$. Stationary but autocorrelated $\{(X_t, Z_t)\}_{t=1}^T$ independent of $\{\varepsilon_t\}_{t=1}^T$ yield that unconditionally, $T^{1/2}(\hat{\theta}_2 - \theta_2) \Rightarrow \mathcal{N}(0, V_2)$, with V_2 in general different from the limit of White's estimator. Finally, under weak exogeneity where, for instance, $X_t = Y_{t-1}$ and $Z_t = Y_{t-2}$, $\hat{\theta}_2$ is inconsistent.

random $q \times 1$ vectors $\{y_{T,t}\}_{t=1}^{T}$, defined on a probability space $(\Omega, \mathfrak{S}, P)$, on which also all following random elements are defined. A triangular array construction for the data is necessary to accommodate the partial instability in the parameter θ.

The GMM population moment condition is embodied in the known, integrable function $g : \mathbb{R}^q \times \Theta \mapsto \mathbb{R}^p$ for $p \geq m$, such that in the stable GMM model, the true parameter θ_0 satisfies $E[g(y_{T,t}, \theta_0)] = 0$ for all $t \leq T$. Let $\{\theta_{T,t}\}_{t=1}^{T} \in \Theta^T$ be the parameter path in the corresponding unstable model such that:

$$E[g(y_{T,t}, \theta_{T,t})] = 0 \quad \text{for all } t \leq T, T \geq 1. \tag{5}$$

For notational convenience, we drop the dependence of $y_{T,t}$ and $\theta_{T,t}$ on T if no confusion arises. Also, let $g_t(\theta)$ be $g(y_t, \theta)$. All limits are taken as $T \to \infty$. We write "$\xrightarrow{p}$" for convergence in probability (in P), "$\Rightarrow$" for weak convergence of the underlying probability measures, $[\cdot]$ denotes the greatest lesser integer function, and $\|\cdot\|$ is the spectral matrix norm. The delimiters of integrals are 0, and 1, if not indicated otherwise.

We analyse the asymptotic properties of the usual GMM estimator $\hat{\theta}$, defined as:

$$\left[T^{-1} \sum_{t=1}^{T} g_t(\hat{\theta}) \right]' Q_T \left[T^{-1} \sum_{t=1}^{T} g_t(\hat{\theta}) \right] = \inf_{\theta \in \Theta} \left[T^{-1} \sum_{t=1}^{T} g_t(\theta) \right]' Q_T \left[T^{-1} \sum_{t=1}^{T} g_t(\theta) \right], \tag{6}$$

where Q_T is a sequence of (possibly random) $p \times p$ positive definite matrices. Denote by $G_t(\theta) = G_{T,t}(y_{T,t}, \theta)$, the $p \times m$ matrix of the partial derivatives $\partial g(y_{T,t}, \theta) / \partial \theta'$ (if it exists).

We impose the following high-level condition.

Condition 1. *The unstable GMM model satisfies the following:*

(i) $T^{1/2}(\theta_t - \theta_0) = f(t/T) \; \forall t \leq T, T \geq 1$ *for some non-stochastic, bounded, and piece-wise continuous function* $f : [0, 1] \mapsto \mathbb{R}^m$ *with at most a finite number of discontinuities, and left and right limits everywhere.*

(ii) *In some neighbourhood* Θ_0 *of* θ_0, $g_t(\theta)$ *is differentiable in* θ *a.s. for* $t \leq T, T \geq 1$.

(iii) $T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t) \Rightarrow \mathcal{N}(0, V)$ *for some positive definite* $p \times p$ *matrix* V.

(iv) $\hat{\theta} \xrightarrow{p} \theta_0$.

(v) $Q_T \xrightarrow{p} Q_0$ *for some positive definite matrix* Q_0, *and there exist positive definite* $p \times p$ *matrices* $\hat{V}_T$ *such that* $\hat{V}_T \xrightarrow{p} V$.

(vi) $T^{-1} \sum_{t=1}^{T} \|G_t(\theta_0)\| = O_p(1)$, $T^{-1} \sup_{t \leq T} \|G_t(\theta_0)\| \xrightarrow{p} 0$ *and for any decreasing sequence of neighbourhoods* $\Theta_T \subset \Theta_0$ *of* θ_0 *that converge to* $\{\theta_0\}$, $T^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta_T} \|G_t(\theta) - G_t(\theta_0)\| \xrightarrow{p} 0$.

(vii) *For all* $0 \leq \lambda \leq 1$, $T^{-1} \sum_{t=1}^{[\lambda T]} G_t(\theta_0) \xrightarrow{p} \lambda \Gamma$ *for some full column rank* $p \times m$ *matrix* Γ.

Part (i) of Condition 1 assumes the instability in the parameters to be of order $T^{-1/2}$. This is the neighbourhood in which efficient tests of parameter stability have non-trivial local asymptotic power. The form of the instability is described by the function f. By letting some elements of f to be zero, the GMM model becomes only partially unstable. The main interest of the paper is how to conduct asymptotically valid inference about the stable subset of parameters. The restrictions on the non-zero parts of the function f are quite weak; in particular, note that we do not assume differentiability of f. The conditions on f are sufficient to ensure that f can be uniformly approximated by a sequence of step functions.

The parameter instability is assumed to be non-stochastic, in contrast to, say, Stock and Watson (1996, 1998), Primiceri (2005), and Cogley and Sargent (2005). But under an alternative

assumption of stochastic parameter paths, the following results continue to hold as long as Condition 1 holds conditional on almost all realizations of the path. Such an assumption, of course, restricts the possible dependence between the disturbances of the model and the stochastic parameter path, but it covers the models of exogenous time-varying parameters models popular in applied work, including those cited above. Almost all realizations of a Wiener process on the unit interval are bounded and continuous and hence may serve as functions f as specified in part (i).

The key assumption for the result in this paper is the approximate linearity of $T^{-1} \sum_{t=1}^{[\lambda T]} G_t(\theta_0)$ in λ as imposed in part (vii) (which, given the condition in part (vi), is equivalent to the approximate linearity of $T^{-1} \sum_{t=1}^{[\lambda T]} G_t(\theta_t)$). This assumption entails that averages of $G_t(\theta_0)$ are approximately equal to Γ in all parts of the sample. It is typically justified for globally stationary models, such as stationary vector autoregressive models. Even certain globally non-stationary models, such as a linear regression with stationary regressors but trending disturbance variance, can satisfy this requirement. On the other hand, most models that generate (stochastically or deterministically) trending data fail to satisfy part (vii) of Condition 1, even after scale normalizations that ensure $T^{-1} \sum_{t=1}^{T} G_t(\theta_0) = O_p(1)$.

Part (iii) assumes a multivariate central limit theorem to hold for the scaled sample average of the moment condition, evaluated at the true time-varying parameter. Given the GMM population moment condition (5), this is a natural condition. At the same time, in order to invoke such a central limit theorem, a suitable set of moment and dependence conditions on the random variables $\{g_t(\theta_t)\}_{t=1}^{T}$ needs to be checked in the unstable model, a complication to which we return in Section 2.4.

Parts (iv)–(vii) impose high-level conditions on the asymptotic properties of the unstable GMM model, which would be fairly standard for a *stable* model, that is if f was equal to 0. Part (iv) can usually be justified by the uniform convergence of $\left[T^{-1} \sum_{t=1}^{T} g_t(\theta)\right]' Q_T \left[T^{-1} \sum_{t=1}^{T} g_t(\theta)\right]$ over $\theta \in \Theta$ to a non-stochastic function whose unique minimizer is θ_0. A suitable estimator $\hat{V}_T$ of V, the asymptotic variance of $T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t)$, is typically given by the non-parametric long-run variance estimators of Newey and West (1987) and Andrews (1991). The third assumption in part (vi) controls the average variability of $G_t(\theta)$ as a function of the parameters. It is implied by the more primitive conditions A.2 and A.3 of Andrews (1987). See Gallant and White (1988) and Andrews (1992) for further discussion. Again, for unstable models with non-zero f, these convergences in probability are less standard, and we provide a suitable argument in Section 2.4.

2.2. *Validity of standard GMM inference on stable parameters*

The following main result establishes the asymptotic properties of standard GMM inference that ignores the parameter instability.

Theorem 1.   *Under Condition* 1,

(i) $T^{1/2} \hat{\Sigma}_\theta^{-1/2} (\hat{\theta} - T^{-1} \sum_{t=1}^{T} \theta_t) \Rightarrow \mathcal{N}(0, I_m)$,

(ii) $T^{-1/2} \sum_{t=1}^{T} g_t(\hat{\theta}) \Rightarrow \mathcal{N}(0, (I_p - \Gamma(\Gamma' Q_0 \Gamma)^{-1} \Gamma' Q_0) V (I_p - \Gamma(\Gamma' Q_0 \Gamma)^{-1} \Gamma' Q_0)')$,
 where $\hat{\Sigma}_\theta = (\hat{\Gamma}' Q_T \hat{\Gamma})^{-1} \hat{\Gamma}' Q_T \hat{V}_T Q_T \hat{\Gamma} (\hat{\Gamma}' Q_T \hat{\Gamma})^{-1}$ *and* $\hat{\Gamma} = T^{-1} \sum_{t=1}^{T} G_t(\hat{\theta}) \xrightarrow{p} \Gamma$.
 Furthermore, if in addition, $\sup_{\lambda \in [0,1]} \left\| T^{-1} \sum_{t=1}^{[\lambda T]} G_t(\theta_0) - \lambda \Gamma \right\| \xrightarrow{p} 0$ *and*
 $T^{-1/2} \sum_{t=1}^{[\cdot T]} g_t(\theta_t) \Rightarrow V^{1/2} W(\cdot)$ *with* W *a* $p \times 1$ *standard Wiener process, then*

(iii) $T^{-1/2} \sum_{t=1}^{[\cdot T]} g_t(\hat{\theta}) \Rightarrow \zeta(\cdot)$, *where* $\zeta(\lambda) = V^{1/2} W(\lambda) - \lambda \Gamma(\Gamma' Q_0 \Gamma)^{-1} \Gamma' Q_0 V^{1/2} W(1) + \Gamma(\int_0^\lambda f(l)dl - \lambda \int_0^1 f(l)dl)$.

Part (i) of Theorem 1 shows that standard asymptotically Gaussian inference based on $\hat{\theta}$ and $\hat{\Sigma}_\theta$ remains valid for the stable subset of the parameters (where θ_t is the same for all t and equal to θ_0 in the corresponding row): for the stable subset, the conventional GMM estimator is asymptotically unbiased and Gaussian. The Wald statistics involving only stable parameters are asymptotically chi-squared under the null hypothesis and have the same non-centrality parameter under local alternatives as the corresponding fully stable model. It is immediate from Condition 1, parts (i) and (iv), that the GMM estimator $\hat{\theta}$ is consistent for the average parameter value, $\left\| \hat{\theta} - T^{-1}\sum_{t=1}^{T}\theta_t \right\| \xrightarrow{P} 0$. Part (i) of Theorem 1 shows how to conduct asymptotically valid inference about this average. In most applications, however, the average of a time-varying parameter does not have a structural interpretation.

To see why the partial instability does not spill over to the estimators of the stable subset of parameters, consider the following first-order Taylor expansion of the first-order condition for equation (6):

$$
0 = \hat{\Gamma}' Q_T T^{-1/2} \sum_{t=1}^{T} g_t(\hat{\theta})
$$

$$
= \hat{\Gamma}' Q_T T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t) + \hat{\Gamma}' Q_T (T^{-1} \sum_{t=1}^{T} \tilde{G}_t) T^{1/2}(\hat{\theta} - \theta_0)
$$

$$
- \hat{\Gamma}' Q_T T^{-1} \sum_{t=1}^{T} \tilde{G}_t T^{1/2}(\theta_t - \theta_0), \tag{7}
$$

where the j-th row of $\tilde{G}_t$ is the j-th row of G_t evaluated at some $\tilde{\theta}_{t,j}$ that lies on the line segment between θ_t and $\hat{\theta}$. Standard arguments imply that under Condition 1, $T^{-1}\sum_{t=1}^{T}\tilde{G}_t \xrightarrow{P} \Gamma$. The main insight concerns the term $T^{-1}\sum_{t=1}^{T}\tilde{G}_t T^{1/2}(\theta_t - \theta_0) = T^{-1}\sum_{t=1}^{T}\tilde{G}_t f(t/T)$. This is a weighted average of the columns of $\{\tilde{G}_t\}_{t=1}^{T}$, with weights $\{f(t/T)\}_{t=1}^{T}$. If the averages of $G_t(\theta_0)$ (and hence $\tilde{G}_t$) are approximately equal to Γ in all parts of the sample, as assumed in Condition 1, part (vii), then the weighted average is approximately the simple average $\times$ the average weight: $T^{-1}\sum_{t=1}^{T}\tilde{G}_t f(t/T) \xrightarrow{P} \lim_{T\to\infty}\Gamma T^{-1}\sum_{t=1}^{T} f(t/T)$. In the context of deriving the asymptotic local power of stability tests, similar results were established in Ploberger, Krämer and Kontrus (1989); Andrews (1993); and Sowell (1996), also see Stock and Watson (1998). Theorem 1, part (i), now follows from rearranging equation (7) and taking limits, revealing the relevance of this result for conducting asymptotically valid inference in partially unstable models.

As a consequence of part (ii) of Theorem 1, the overidentification test by Hansen (1982) remains asymptotically chi-squared with $p - m$ degrees of freedom, even in the unstable model. The overidentification test has no power against the alternative of (locally) time-varying parameters—this result was obtained by Ghysels and Hall (1990) for a single break and is implied by Sowell's (1996) asymptotic decomposition of the sample moment condition, also see Newey (1985) and Hall and Sen (1999). Therefore, when conducting inference about stable parameters in a partially unstable model as described in Condition 1, rejection by the overidentification test cannot be explained by the partial instability. As usual, it still indicates incorrect moment conditions.

2.3. *Stability tests in partially unstable models*

Part (iii) of Theorem 1 requires the strengthening of Condition 1, part (iii), to a functional central limit theorem to hold for the partial sums of the sample moment conditions evaluated at the true time-varying parameter, and the convergence in Condition 1, part (vii), to be uniform. The result

serves as a basis for understanding the asymptotic local power of a wide range of parameter stability tests. The statistics analysed in Nyblom (1989), Sowell (1996) and Elliott and Müller (2006), as well as the LM versions of the tests derived in Andrews (1993) and Andrews and Ploberger (1994) can be written as functions of $T^{-1/2}\sum_{t=1}^{[\cdot T]}g_t(\hat{\theta})$. Of special interest here are the properties of stability tests in partially unstable models. Suppose one is interested in the first $m_0 \leq m$ elements of θ. Let C be the $m \times m_0$ selection matrix $C = [I_{m_0}, 0_{m_0 \times (m-m_0)}]'$, and consider the case of efficient GMM estimation, so that $Q_T = \hat{V}_T^{-1}$. One might invoke the analysis of Sowell (1996), who derives tests of:

$$H_0 : \theta_t \text{ is constant in } t \quad \text{against} \quad H_1 : \theta_t \text{ depends on } t, \tag{8}$$

that efficiently discriminate between the *limiting* distribution of $T^{-1/2}\sum_{t=1}^{[\cdot T]}g_t(\hat{\theta})$ under the null and local alternatives of equation (8). Specifically, Sowell's Corollary 2 shows that the limit random process:

$$J(\lambda) = \Sigma_\theta^{-1/2}(W_m(\lambda) - \lambda W_m(1)) + \Sigma_\theta^{-1}\left(\int_0^\lambda f(l)dl - \lambda \int_0^1 f(l)dl\right),$$

of $\hat{\Gamma}'\hat{V}_T^{-1}T^{-1/2}\sum_{t=1}^{[\cdot T]}g_t(\hat{\theta}) \Rightarrow J(\cdot)$ is asymptotically sufficient for the local alternative f, where $\Sigma_\theta = (\Gamma'V^{-1}\Gamma)^{-1}$ and W_m is a $m \times 1$ standard Wiener process, so that $W_m(\lambda) - \lambda W_m(1)$ is a $m \times 1$ Brownian Bridge. Specializing his Corollary 2 further, one obtains that tests of the hypothesis in (8) that maximize power against alternatives where only the first m_0 elements of θ are time varying may be based on functionals of:

$$(C'\hat{\Sigma}_\theta^{-1}C)^{-1/2}C'\hat{\Gamma}'\hat{V}_T^{-1}T^{-1/2}\sum_{t=1}^{[\cdot T]}g_t(\hat{\theta})$$

$$\Rightarrow W_{m_0}(\cdot) - \cdot W_{m_0}(1) + (C'\Sigma_\theta^{-1}C)^{-1/2}C'\Sigma_\theta^{-1}\left(\int_0^\cdot f(l)dl - \cdot \int_0^1 f(l)dl\right), \tag{9}$$

where $W_{m_0} = (C'\Sigma_\theta^{-1}C)^{-1/2}C'\Sigma_\theta^{-1/2}W_m$. In general, as long as Σ_θ is not block diagonal, the asymptotic distribution (9) depends on whether or not the last $m - m_0$ elements in f are zero. The asymptotic null distribution of the usual tests for instability in the first m_0 elements of θ are thus typically affected by instabilities in other parameters, as long as the parameter estimators are not asymptotically uncorrelated. In other words, these tests are not in general valid tests of:

$$H_0 : C'\theta_t \text{ is constant in } t \quad \text{against} \quad H_1 : C'\theta_t \text{ depends on } t, \tag{10}$$

which allows for local instabilities of the last $m - m_0$ parameters in θ under the null hypothesis.

As a solution to this problem, consider the class of modified test statistics that are functions of:

$$(C'\hat{\Sigma}_\theta C)^{-1/2}C'\hat{\Sigma}_\theta\hat{\Gamma}'\hat{V}_T^{-1}T^{-1/2}\sum_{t=1}^{[\cdot T]}g_t(\hat{\theta}) \Rightarrow (C'\Sigma_\theta C)^{-1/2}C'\Sigma_\theta J(\cdot)$$

$$= \tilde{W}_{m_0}(\cdot) - \cdot \tilde{W}_{m_0}(1) + (C'\Sigma_\theta C)^{-1/2}C'\left(\int_0^\cdot f(l)dl - \cdot \int_0^1 f(l)dl\right), \tag{11}$$

where $\tilde{W}_{m_0} = (C'\Sigma_\theta C)^{-1/2} C' \Sigma_\theta^{1/2} W_m$. The asymptotic null distribution of stability tests based on equation (11) does not depend on the local instabilities in the last $m - m_0$ elements of θ because $C'f$ is equal to 0 whenever the first m_0 elements of f are. More formally, the R.H.S. of equation (11) is a maximal invariant of J for the group of transformations $J \mapsto J + \Sigma_\theta^{-1} (0_{1 \times m_0}, \tilde{F}')'$ for continuous functions $\tilde{F} : [0, 1] \mapsto \mathbb{R}^{m - m_0}$ satisfying $\tilde{F}(0) = \tilde{F}(1) = 0$. This group corresponds to the transformations induced by local instabilities in the last $m - m_0$ elements of θ, so that efficient tests based on equation (11) yield best invariant tests of the hypothesis in (8) based on J.[3]

When one applies the same type of test statistic to equations (9) and (11), such as the functional corresponding to Nyblom's (1989) statistic $N(\psi(\cdot)) = \int_0^1 \psi(\lambda)' \psi(\lambda) d\lambda$, where $\psi(\cdot)$ is the L.H.S. of equations (9) and (11), one obtains the same asymptotic distribution when all parameters are stable and thus the same critical value. But in contrast to equation (9), under the null hypothesis (10) of stability of the first m_0 elements of θ, the second summand in equation (11) is equal to zero, independent of the last $m - m_0$ elements of f. Therefore, as long as all potential instabilities are local, one might only test the stability of those parameters that one is actually interested in, and the result of tests based on equation (11) will not be affected by instabilities in the non-tested parameters.[4] If a stability test based on a functional of (11) rejects, it indicates that the presumably stable subset of parameters is not stable after all.

If it is known for sure that the last $m - m_0$ parameters are stable, however, tests based on equation (11) typically have lower power than tests based on equation (9). By the formula for the inverse of a partitioned matrix, $C'\Sigma_\theta^{-1} C - (C'\Sigma_\theta C)^{-1}$ is positive semi-definite, and zero only if Σ_θ is block diagonal. The "signal-to-noise ratio" against alternatives of the form $f = Cg$ for some function $g : [0, 1] \mapsto \mathbb{R}^{m_0}$ in equation (9) is $(C'\Sigma_\theta^{-1} C)^{-1/2} C' \Sigma_\theta^{-1} C = (C'\Sigma_\theta^{-1} C)^{1/2}$, which is larger than the corresponding ratio $(C'\Sigma_\theta C)^{-1/2} C'C = (C'\Sigma_\theta C)^{-1/2}$ in equation (11). For tests that seek to detect potential instabilities in all parameters, that is $C = I_m$, equation (11) reduces to equation (9).

In summary, for a partially unstable GMM model under Condition 1, standard asymptotically Gaussian GMM inference about the stable subset of parameters remains valid. Also, rejection of the overidentification test continues to indicate mistaken moment conditions. The rejection probability of usual stability tests for a subset of parameters, in contrast, is typically affected by instabilities in the non-tested parameters. As a solution, we suggest basing inference on a class of modified statistics that are functions of equation (11), whose asymptotic rejection probabilities are a function of the stability of the parameters under consideration only.

2.4. *Justifying Condition* 1 *via contiguity*

As noted in Section 2.1, the assumptions in parts (iii)–(vii) of Condition 1 are fairly standard for stable GMM models. The analysis of unstable models is complicated by the fact that parameter instability typically leads to non-stationary data and potentially complicated interactions between the time-varying parameters and the data generating process (think of regression models with lagged dependent variables with time-varying coefficients). One way to address these complications is to restrict the possible interactions. Ploberger *et al.* (1989) only consider regression models with strictly exogenous regressors. Sowell (1996) assumes that both the stable and

---

3. Müller (2007) provides a sense in which this procedure is asymptotically efficient for the original GMM problem, rather than only for the limiting process J.
    4. One might worry about the local nature of the robustness against instabilities in the last $m - m_0$ elements of θ of this approach. Formally, one might resolve this issue by never rejecting H_0 in (10) whenever the p value of a parameter stability test for the last $m - m_0$ elements in θ is smaller than some threshold $\varepsilon > 0$, similar to the discussion in the Introduction.

the unstable models generate stationary data. In the context of an unstable regression, Stock and Watson (1998) rule out lagged dependent variables.

It might be possible to justify Condition 1 directly by imposing primitive conditions on the unstable model similar to those in Andrews (1993) (see Ghysels, Guay and Hall, 1997, and Hall and Sen, 1999, for additional results based on these assumptions). In Andrews' (1993) analysis of the local asymptotic power of stability tests, $\{g_t(\theta_0)\}_{t=1}^T$ is assumed to be near-epoch dependent with time-varying mean and finite higher moments. Such conditions allow for a rich set of unstable models, including regression models with only weakly exogenous regressors. At the same time, given the highly technical nature of these primitive assumptions, for any given model, it might not be much harder to establish the high-level Condition 1 from first principles. Also, Andrews (1993) does not provide a discussion of the consistency of the long-run variance estimator $\hat{V}_T$ in the unstable model.

We hence refrain from further discussing primitive conditions on the data and the function g that imply Condition 1 directly. Rather, we now discuss conditions on the likelihood of stable models that imply Condition 1, parts (iii)–(vii), to hold in the unstable model whenever they hold in the corresponding stable model. This indirect reasoning circumvents much of the difficulty of establishing Condition 1 in (locally) unstable models.

The difference between the unstable model and the corresponding stable model is the presence of time-varying parameters, whose time variation is only big enough to be detectable with some (possibly high) probability. Even efficient GMM-based tests for parameter stability cannot discriminate between the stable and the unstable models consistently. But this suggests that no statistic can be of a different probabilistic order in the unstable model than in the stable model. This in turn implies Condition 1, parts (iv)–(vii), to be true in the unstable GMM model whenever they hold in the corresponding stable GMM model (*i.e.* when $f = 0$). The formal argument relies on the concept of *contiguity*.

Definition 1.   Let P_T^1 and P_T^2 be two sequences of probability measures on a measurable space $(\Omega_T, \mathcal{F}_T)$. The sequence P_T^1 is called contiguous to P_T^2 if $X_T \overset{p}{\to} 0$ under P_T^2 implies $X_T \overset{p}{\to} 0$ under P_T^1 for all random variables $X_T : \Omega_T \mapsto \mathbb{R}$.

See, for instance, chapter 6 of van der Vaart (1998) for further discussion and references.

To make the heuristic reasoning rigorous, we thus need to impose some regularity conditions on the generating process of the data $\{y_{T,t}\}_{t=1}^T$ to ensure that the unstable model is contiguous to the stable model. Assume that the difference between the density of the stable and unstable models can be described by the evolution of the $k \times 1$ parameter β, $k \geq m$, such that for all $s \leq T$, the density of $\{y_{T,t}\}_{t=1}^s$ (with respect to some sigma finite measure) is given by $\prod_{t=1}^s f_{T,t}(y_{T,t}, y_{T,t-1}, \ldots, y_{T,1}; \beta_{T,t})$ when β takes on the value $\beta_{T,t}$ at date t. With $k > m$, this allows the instability in the likelihood to go beyond the instability in the GMM parameter θ. Denote by $l_{T,t}(\beta) = \ln f_{T,t}(y_{T,t}, y_{T,t-1}, \ldots, y_{T,1}; \beta)$ the contribution to the log likelihood of the density at date t, the scores $s_{T,t}(\beta) = \partial l_{T,t}(\beta)/\partial\beta$ and the Hessians $h_{T,t}(\beta) = \partial s_{T,t}(\beta)/\partial\beta'$. Let $\mathfrak{F}_{T,t}$ be the σ field generated by $\{y_{T,s}\}_{s=1}^t$, and $\mathfrak{F}_{T,0}$ be the trivial σ field. We again omit the dependence on T of $\beta_{T,t}$, $s_{T,t}$, $h_{T,t}$, and $\mathfrak{F}_{T,t}$ for notational simplicity. Also, we refer to the model with density $\prod_{t=1}^T f_{T,t}(y_{T,t}, y_{T,t-1}, \ldots, y_{T,1}; \beta_0)$ as the "stable model".

Condition 2.

(i) *The unstable parameter vector β_t satisfies $T^{1/2}(\beta_t - \beta_0) = B(t/T)$ for some bounded and piece-wise continuous vector function $B : [0, 1] \mapsto \mathbb{R}^k$ with at most a finite number of discontinuities, and left and right limits everywhere.*

(ii) *In some neighbourhood $\mathcal{B}_0$ of β_0, $l_t(\beta)$ is twice differentiable a.s. with respect to β for $t = 1, \ldots, T$.*

Furthermore, in the stable model,

(iii) *$\{s_t(\beta_0), \mathfrak{F}_t\}$ is a square-integrable martingale difference array with $T^{-1} \sum_{t=1}^{[\lambda T]} E[s_t(\beta_0)s_t (\beta_0)' \mid \mathfrak{F}_{t-1}] \overset{p}{\to} \int_0^\lambda \Upsilon(l)dl$ for all $0 \leq \lambda \leq 1$ and some non-stochastic bounded Riemann integrable matrix function $\Upsilon : [0, 1] \mapsto \mathbb{R}^{k \times k}$, $T^{-1} \sup_{t \leq T} \left\| E[s_t(\beta_0)s_t(\beta_0)' \mid \mathfrak{F}_{t-1}] \right\| \overset{p}{\to} 0$ and there exists $\nu > 0$ such that $T^{-1} \sum_{t=1}^T E[\|s_t(\beta_0)\|^{2+\nu} \mid \mathfrak{F}_{t-1}] = O_p(1)$.*

(iv) *$T^{-1} \sum_{t=1}^T \|h_t(\beta_0)\| = O_p(1)$, $T^{-1} \sup_{t \leq T} \|h_t(\beta_0)\| \overset{p}{\to} 0$ and for any decreasing sequence of neighbourhoods $\mathcal{B}_T \subset \mathcal{B}_0$ of β_0 that converge to $\{\beta_0\}$, $T^{-1} \sum_{t=1}^T \sup_{\beta \in \mathcal{B}_T} \|h_t(\beta) - h_t(\beta_0)\| \overset{p}{\to} 0$.*

(v) *For all $0 \leq \lambda \leq 1$, $T^{-1} \sum_{t=1}^{[\lambda T]} h_t(\beta_0) \overset{p}{\to} - \int_0^\lambda \Upsilon(l)dl$.*

Part (i) makes the same assumption on the form of the instability in β as Condition 1, part (i), does on θ. Parts (iii)–(v) are weak regularity conditions on the likelihood of the stable model; see, for instance, Phillips and Ploberger (1996) for a similar set of assumptions. When integration and differentiation can be exchanged and the relevant conditional moments exist, $\{s_t(\beta_0), \mathfrak{F}_t\}$ and $\{s_t(\beta_0)s_t(\beta_0)' + h_t(\beta_0), \mathfrak{F}_t\}$ are martingale difference arrays by construction (see Hall and Heyde, 1980, chapter 6.2). The matrix function Υ represents the average rate of (conditional) information accrual on the time scale of the the sample fraction. For stationary stable models, Υ is constant and equal to the probability limit of $(-T^{-1} \sum_{t=1}^T h_t(\beta_0))$ and $T^{-1} \sum_{t=1}^T E[s_t(\beta_0)s_t(\beta_0)' \mid \mathfrak{F}_{t-1}]$. The point-wise convergences in λ in parts (iii) and (v) are then fulfilled automatically.

Lemma 1. *Under Condition* 2, *the unstable model is contiguous to the stable model. In particular, if a stable GMM model satisfies Condition* 1, *parts* (iv)–(vii), *and Condition* 2, *then Condition* 1, *parts* (iv)–(vii), *also holds under the unstable model.*

Lemma 1 formally states the possibility of obtaining Condition 1, parts (iv)–(vii), by making assumptions only on the stable GMM model. As argued above, Condition 1, parts (iv)–(vii), is quite standard under stability. Note that one does not need to know the likelihood structure of the data to take advantage of this reasoning, as long as one is willing to assume Condition 2. In a general GMM setup, Condition 2 plays the role of a regularity condition, akin to more familiar mixing or moment conditions.

Under an alternative assumption of a stochastic parameter path, as discussed in Section 2.1, the argument based on contiguity may still be applied, as long as Condition 2 holds conditional on almost all realizations of B. See the Appendix for details.

While contiguity implies that all $o_p(1)$ approximations of the stable model remain asymptotically accurate in the unstable model, it does not in itself justify Condition 1, part (iii), the weak convergence of the average sample moment condition to a multivariate normal. At the same time, some primitive conditions of (functional) central limit theorems take the form of convergences in probability. To establish those in the unstable model, it suffices to show that they hold in the stable model and to then invoke contiguity. As an example, consider the case where the moment condition evaluated at the truth $g_{T,t}(\theta_{T,t})$ is a martingale difference array with respect to the sigma fields $\mathfrak{G}_{T,t}$, where $g_{T,s}(\theta_{T,s})$ is measurable with respect to $\mathfrak{G}_{T,t}$ for all $s < t$. Dropping again the dependence on T for simplicity, we can verify the conditions given in McLeish (1974) and establish the following lemma.

Lemma 2. *If in the unstable model, $\{g_t(\theta_t), \mathfrak{G}_t\}_{t=1}^T$ is a martingale difference array and there exists $\varsigma > 0$ such that $T^{-1} \sum_{t=1}^T E[\|g_t(\theta_t)\|^{2+\varsigma} \mid \mathfrak{G}_{t-1}] = O_p(1)$, and in the stable model, Condition 1, parts* (i),(ii),(vi), *and* (vii), *hold, $T^{-1/2} \sup_{t \leq T} \|g_t(\theta_0)\| \xrightarrow{p} 0$ and $T^{-1} \sum_{t=1}^T g_t(\theta_0) g_t(\theta_0)' \xrightarrow{p} V$ for some positive definite $p \times p$ matrix V, then under Condition 2, $T^{-1/2} \sum_{t=1}^T g_t(\theta_t) \Rightarrow \mathcal{N}(0, V)$ in the unstable model. Furthermore, if in addition $T^{-1} \sum_{t=1}^{[\lambda T]} g_t(\theta_0) g_t(\theta_0)' \xrightarrow{p} \lambda V$ for all $0 \leq \lambda \leq 1$ in the stable model, then $T^{-1/2} \sum_{t=1}^{[\cdot T]} g_t(\theta_t) \Rightarrow V^{1/2} W(\cdot)$ in the unstable model, where W is a $p \times 1$ standard Wiener process.*

To apply Lemma 2, the only condition that needs to be verified in the unstable model is that $\{g_t(\theta_t), \mathfrak{G}_t\}_{t=1}^T$ is a martingale difference array with slightly more than two conditional moments, which are bounded in probability on average. This is often further facilitated by contiguity. Suppose $g_t(\theta_t)$ is of the form $x_{t-1} \varepsilon_t$ in the unstable model, with x_t measurable with respect to $\mathfrak{G}_t$ and $\sup_t E[\|\varepsilon_t\|^{2+\varsigma} \mid \mathfrak{G}_{t-1}] \leq \bar{M}_\varepsilon$ a.s under the unstable model. Then, $T^{-1} \sum_{t=1}^T E[\|g_t(\theta_t)\|^{2+\varsigma} \mid \mathfrak{G}_{t-1}] \leq \bar{M}_\varepsilon T^{-1} \sum_{t=1}^T \|x_{t-1}\|^{2+\varsigma}$ a.s. in the unstable model, and it suffices to show that $T^{-1} \sum_{t=1}^T \|x_{t-1}\|^{2+\varsigma} = O_p(1)$ in the stable model to conclude by contiguity that it is also $O_p(1)$ in the unstable model.

Interestingly, one can justify Condition 1, part (iii), entirely with assumptions on the stable model when the likelihood can be parametrized in a way such that the moment condition becomes a linear combination of the derivatives of the log likelihood. The leading case for this is, of course, maximum likelihood estimation, although it also covers instances where only a subset of the likelihood derivatives is exploited as moment conditions. The proof of the following lemma relies heavily on LeCam's third lemma (see van der Vaart, 1998, p. 90), an asymptotic change of measure from the stable to the unstable model.

Lemma 3. *If Condition 2 holds and $\left\| T^{-1/2} \sum_{t=1}^T g_t(\theta_0) - T^{-1/2} F' \sum_{t=1}^T s_t(\beta_0) \right\| \xrightarrow{p} 0$ under the stable model for some $k \times p$ matrix F, then $T^{-1/2} \sum_{t=1}^T g_t(\theta_0) \Rightarrow \mathcal{N}(0, V)$ in the stable model and $T^{-1/2} \sum_{t=1}^T g_t(\theta_t) \Rightarrow \mathcal{N}(0, V)$ in the unstable model, where $V = F' \int \Upsilon(s) ds\, F$.*

To sum up, a reasoning via contiguity justifies the high-level Condition 1 for the unstable model mostly by reference to the corresponding stable model: whenever a stable model satisfies Conditions 1 and 2, then Condition 1, parts (iv)–(vii), also hold under the unstable model. In general, Condition 1, part (iii), under the unstable model requires an additional argument, but contiguity either simplifies the application of an appropriate central limit theorem (Lemma 2) or, in the special context of Lemma 3, is also implied by contiguity whenever Condition 1, part (iii), holds in the stable model. Theorem 1 thus holds for a wide range of data generating processes, including regression models with lagged endogenous variables and models with additional local time variation in unmodelled parameters.

3. MONTE CARLO RESULTS

The results of the last section show that usual GMM inference about a stable subset of parameters remains asymptotically valid under certain conditions. The two main assumptions are that (i) the instabilities are local in the sense that even efficient stability tests do not detect them with probability 1 and (ii) the derivative of the moment sample condition has approximately equal averages in all parts of the sample. This section explores the accuracy of this asymptotic result in small samples by two Monte Carlo experiments. The first experiment concerns a simple time series OLS regression, and we explore the quantitative implications of assumption (ii) by considering

null rejection probabilities as a function of the persistence in the regressors. The second experiment concerns GMM inference for the structural parameters in the NKPC example introduced in the Introduction. We calibrate the data generating process to quarterly U.S. data and explore the range of magnitudes of the instability for which our results remain useful approximations.

3.1. *OLS regression*

The first experiment considers the linear regression:

$$Y_t = X_t\theta_{1,t} + Z_t\theta_2 + \theta_3 + \varepsilon_t, t = 1, \ldots, T, \tag{12}$$

where $\varepsilon_t \sim$ i.i.d. $\mathcal{N}(0, 1)$ and $(X_t, Z_t)'$ is a zero-mean stationary Gaussian VAR(1) independent of $\{\varepsilon_t\}$ with coefficient matrix rI_2, $EX_t^2 = EZ_t^2 = 1$ and $E[X_tZ_t] = \rho_{XZ}$. Let $R_t = (X_t, Z_t, 1)'$, and denote by $\hat{\varepsilon}_t$ the OLS residuals of regression (12). This is an exactly identified GMM problem, where the derivative of the moment condition equals $G_t(\theta) = R_tR_t'$, so that $\hat{\Gamma} = T^{-1}\sum_{t=1}^T R_tR_t'$ and, for heteroskedasticity robust inference, we set $\hat{V}_T = T^{-1}\sum_{t=1}^T R_tR_t'\hat{\varepsilon}_t^2$. The parameter r governs the degree of autoregressive persistence in the non-constant regressors, and our asymptotic results formally apply when $|r| < 1$.

We base tests for the presence of an instability on analogues of Nyblom's (1989) statistic. Let C be a $3 \times m_0$, $m_0 \leq 3$ matrix, which is constructed of those columns of I_3 that correspond to the coefficients whose stability is to be tested. For instance, to test the stability of θ_2, $C = (0, 1, 0)'$. With $\hat{\Sigma}_\theta = (\hat{\Gamma}'\hat{V}_T^{-1}\hat{\Gamma})^{-1}$, the non-modified Nyblom statistic based on equation (9) is then given by:[5]

$$N = T^{-2}\sum_{s=1}^T \left(C'\hat{\Gamma}'\hat{V}_T^{-1}\sum_{t=1}^s R_t\hat{\varepsilon}_t\right)' \left(C'\hat{\Sigma}_\theta^{-1}C\right)^{-1} \left(C'\hat{\Gamma}'\hat{V}_T^{-1}\sum_{t=1}^s R_t\hat{\varepsilon}_t\right), \tag{13}$$

and the modified Nyblom statistic based on equation (11) is:

$$M = T^{-2}\sum_{s=1}^T \left(C'\hat{\Sigma}_\theta\hat{\Gamma}'\hat{V}_T^{-1}\sum_{t=1}^s R_t\hat{\varepsilon}_t\right)' \left(C'\hat{\Sigma}_\theta C\right)^{-1} \left(C'\hat{\Sigma}_\theta\hat{\Gamma}'\hat{V}_T^{-1}\sum_{t=1}^s R_t\hat{\varepsilon}_t\right). \tag{14}$$

By Theorem 1, part (iii), under the null hypothesis of all coefficients being constant, the asymptotic distribution of both N and M is as tabulated in Nyblom (1989).

We consider two forms of instability in θ_1: a "break" in the middle of the sample, $\theta_{1,t} = hT^{-1/2}\mathbf{1}[t > T/2]$, and a Gaussian "random walk", $\theta_{1,t} = hT^{-1/2}W(t/T)$, where W is a standard Wiener process independent of $\{\varepsilon_t, R_t\}_{t=1}^T$. Small instabilities (denoted as "sm" in Table 1) correspond to $h = 5$ and $h = 8$ in the single break case and the random walk case, respectively; large instabilities (denoted as "lg" in Table 1) correspond to $h = 10$ and $h = 16$. We set the sample size $T = 100$ and, as a benchmark, $r = 0.5$. Table 1 reports empirical rejection probabilities of heteroskedasticity robust two-sided t tests on θ_1 and θ_2 (t_1 and t_2) under the null hypothesis, of the usual Nyblom statistics (13) for the constancy of all three coefficients (N_{all}) and of θ_1 and θ_2 (N_1 and N_2), and of the modified Nyblom statistics (14) for the constancy of the coefficients θ_1 and θ_2 (M_1 and M_2). The number of replications is 50,000. All tests are based on 5% nominal-level asymptotic critical values. When $\theta_{1,t}$ is time varying, the "true" value of θ_1 is set

---

5. This version of the heteroskedasticity robust Nyblom's (1989) statistic differs from what is suggested in Hansen (1990) and often employed in practice, that is $T^{-1}\sum_{s=1}^T \left(C'\sum_{t=1}^s R_t\hat{\varepsilon}_t\right)' \left(T^{-1}\sum_{t=1}^T C'R_tR_t'C\hat{\varepsilon}_t^2\right)^{-1}\left(C'\sum_{t=1}^s R_t\hat{\varepsilon}_t\right)$. The optimality result of Sowell (1996) discussed above implies that in the presence of heteroskedasticity, equation (13) is the more powerful statistic, at least asymptotically.

TABLE 1

Rejection probabilities of 5% nominal tests, $T = 100$ and $r = 0.5$

	Break							Random walk						
h	t_1	t_2	N_{all}	N_1	N_2	M_1	M_2	t_1	t_2	N_{all}	N_1	N_2	M_1	M_2
							$\rho_{XZ} = 0$							
0	6·4	6·6	4·3	4·7	4·6	4·6	4·6	6·4	6·6	4·3	4·7	4·6	4·6	4·6
sm	6·8	6·9	32·1	50·2	4·4	49·8	4·4	6·9	7·3	33·6	43·1	4·8	43·0	4·9
lg	7·6	7·9	87·0	96·1	4·1	96·1	3·7	7·4	8·4	59·5	67·5	4·8	68·0	5·4
							$\rho_{XZ} = 0.5$							
sm	6·7	6·9	32·1	50·2	14·3	39·3	4·4	6·8	7·3	33·6	43·1	14·5	36·7	4·9
lg	7·6	7·9	87·0	96·1	30·7	89·7	3·7	7·7	8·4	59·5	67·5	23·1	62·6	5·4
							$\rho_{XZ} = 0.9$							
sm	6·9	6·9	32·1	50·2	41·1	12·9	4·4	7·1	7·3	33·6	43·1	36·9	15·0	4·9
lg	7·9	7·9	87·0	96·1	87·7	35·0	3·7	8·3	8·4	59·5	67·5	60·0	32·1	5·4

to $T^{-1}\sum_{t=1}^{T}\theta_{1,t}$ in the computation of t_1. By inspection of equations (13) and (14), note that M_2 and N_1 are invariant to the transformation $\{R_t\} \mapsto \{AR_t\}$ for any invertible lower triangular matrix A. This transformation leaves the regressor with the time-varying parameter unchanged up to scale, so that the results for M_2 and N_1 do not depend on the value of ρ_{XZ} even in the unstable model, and this also holds for t_2 by the same argument. Similarly, in the stable model with $h = 0$, the distribution of all tests does not depend on ρ_{XZ}.

The empirical rejection probability of the t test on the stable coefficient θ_2 is very little affected by the instability in θ_1, as predicted by Theorem 1, part (i). This remains true even for instabilities that are large enough to be detected by tests with high probability—for the "large" instability, the p value of N_{all} is smaller than 0·1% for more than one in four realizations. The magnitudes of the instabilities considered here are very large by empirical standards. Cogley and Sargent (2005) find instabilities in parameters of monetary VARs that they consider "substantial" from an economic point of view, but which are detected by 5% nominal-level Nyblom statistics less than 25% of the time. Stock and Watson (1996) reject the stability of the seven parameters describing univariate AR(6) models for 40 of 76 U.S. postwar macro-economic time series on the 10% level using Andrews' (1993) supF statistic. But based on the method by Stock and Watson (1998) for obtaining median unbiased estimates for the magnitude h of a random walk instability by inverting the supF test statistic, the largest estimate of h for these 76 models is less than 12. Similarly, in Ghysel's (1998) application in asset pricing, he mostly rejects the stability of two-parameter and three-parameter versions of a conditional consumption-based capital asset pricing model for 12 industry and 10 size-sorted portfolios, using seven different instruments, and often on the 1% significance level. But his test statistics imply median unbiased estimates of h that are always smaller than 11. What is more, for $T = 250$, one needs to double the magnitude of the instabilities to obtain roughly similar size distortions as reported in Table 1. All this suggests that the results of this paper are of empirical relevance for many parameter instabilities that one might encounter in a financial or macro-economic application.

Also, as implied by Theorem 1, part (i), the t test of θ_1 using the pseudo-true value $T^{-1}\sum_{t=1}^{T}\theta_{1,t}$ has a rejection probability close to the nominal level. The usual Nyblom statistic for the stability of θ_2, N_2, is strongly affected by the instability in θ_1 when $\rho_{XZ} \neq 0$, in contrast to the modified statistic M_2. Comparing the power of M_1 and N_1, we find substantial losses in power of the modified statistic only when $\rho_{XZ} = 0.9$.
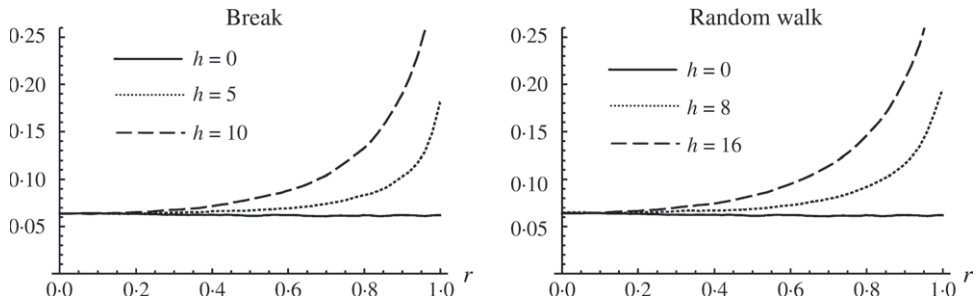
FIGURE 1

Small sample rejection probability of the 5% nominal t test about a stable parameter in an OLS regression as a function of the persistence of the regressors

The results given in Table 1 are based on regressors that follow a VAR(1) process with coefficient $r = 0.5$, so that persistence in the regressors is moderate. Figure 1 depicts the null rejection probability of the t test t_2 of the stable parameter θ_2 as a function of $r \in [0, 1)$. While technically a stationary VAR, for large autoregressive roots $r < 1$, linearity of $T^{-1} \sum_{t=1}^{[\lambda T]} R_t R_t'$ in λ, that is Condition 1, part (vii), is a poor approximation, and the test displays substantial overrejections. As one might expect, unreported simulations show that r needs to be relatively closer to unity to induce overrejections of a similar degree in larger samples. Nevertheless, our results should be applied cautiously for models with persistent data, such as macro-data in levels. We recommend conducting a model-specific Monte Carlo analysis when persistence is an issue.

3.2. *A stylized NKPC model*

The second experiment is based on the NKPC data generating process (3) and (4) of the Introduction. With $\phi < 1$, the unique reduced form of the two-equation system is:

$$\Delta \pi_t = \alpha_1 s_{t-1} + \alpha_2 s_{t-2} + v_t \tag{15}$$
$$s_t = \rho_1 s_{t-1} + \rho_2 s_{t-2} + \xi_t,$$

where $v_t = \varepsilon_t + \gamma \xi_t$ and:

$$\alpha_1 = \frac{\kappa(\rho_1 + \phi \rho_2)}{1 - \phi \rho_1 - \phi^2 \rho_2}, \quad \alpha_2 = \frac{\kappa \rho_2}{1 - \phi \rho_1 - \phi^2 \rho_2}, \quad \text{and} \quad \gamma = \frac{\kappa}{1 - \phi \rho_1 - \phi^2 \rho_2}. \tag{16}$$

We adopt the "anticipated utility" assumption of the learning literature that agents know the true value of the parameters at each period but behave as if the parameters remained constant in the future (*cf.* Kreps, 1998). Under this assumption, time-varying parameters of the model (4) lead to time-varying parameters of the reduced-form parameters in equation (16), with the current values of α_1, α_2, and γ determined by the current values of ρ_1 and ρ_2. Note that due to the interaction via the expected future inflation term, instabilities in ρ_1 and ρ_2 lead to unstable reduced-form parameters α_1 and α_2, even when the Euler equation in (3) is assumed stable throughout.

Leading the first equation in (15) one period and taking expectations conditional on information available at date $t - 1$, the forecasting equation for $\Delta \pi_{t+1}$ becomes $\Delta \pi_{t+1} = (\alpha_1 \rho_1 + \alpha_2) s_{t-1} + \alpha_1 \rho_2 s_{t-2} + v_{t+1} + \alpha_1 \xi_t$. Therefore s_{t-1} and s_{t-2} are the only relevant instruments for the two endogenous regressors $\Delta \pi_{t+1}$ and s_t of equation (3). The two-equation system (3) and
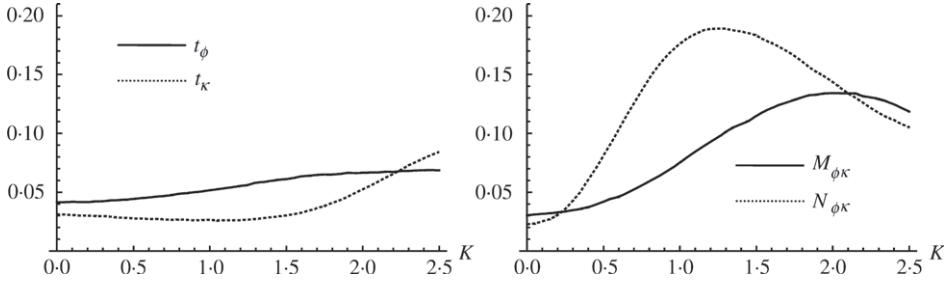
Small sample rejection probabilities of 5% nominal rests in the NKPC data generating process as function of the magnitude of the instability

(4) is therefore exactly identified, and efficient GMM estimation is based on the moment conditions $E[g_t(\theta)] = 0$, where $\theta = (\phi, \kappa, \rho_1, \rho_2)'$ and $g_t(\theta) = ((\Delta\pi_t - \phi\Delta\pi_{t+1} - \kappa s_t)s_{t-1}, (\Delta\pi_t - \phi\Delta\pi_{t+1} - \kappa s_t)s_{t-2}, (s_t - \rho_1 s_{t-1} - \rho_2 s_{t-2})s_{t-1}, (s_t - \rho_1 s_{t-1} - \rho_2 s_{t-2})s_{t-2})'$.

Assuming $(\varepsilon_t, \xi_t)'$ to be i.i.d. mean-zero Gaussian, it is straightforward to see that the stable reduced-form model (15) satisfies Condition 2, parts (iii)–(v), so that Lemma 1 and standard arguments concerning stable GMM models yield Condition 1, parts (iv)–(vii), in an unstable model with parameter instabilities as specified in Condition 2, part (i). Furthermore, under the unstable model, $g_t(\theta_t) = (s_{t-1}(\varepsilon_t - \phi v_{t+1}), s_{t-2}(\varepsilon_t - \phi v_{t+1}), s_{t-1}\xi_t, s_{t-2}\xi_t)'$, so that $g_t(\theta_t)$ is a vector moving average process of order 1, and arguments similar to those applied in the proof of Lemma 2 yield $T^{-1/2}\sum_{t=1}^{[\cdot T]} g_t(\theta_t) \Rightarrow V^{1/2}W(\cdot)$ in the unstable model.

For the Monte Carlo study, the parameter values used in the data generating process are estimated using U.S. quarterly inflation and unemployment series from 1960:1 to 2000:4.[6] For the NKPC (3), we use the full-sample estimates: $\phi = 0.73$ and $\kappa = -0.35$. For the AR(2) process of the unemployment gap (4), the size of the instability used in the Monte Carlo is obtained from split-sample estimation (with a break in the middle of the sample, 1979:4, corresponding to the date of an important change in monetary policy—the start of Chairman Volcker's tenure): the first subsample yields estimates of $\rho_1 = 0.46$ and $\rho_2 = -0.22$, and estimated changes in ρ_1 and ρ_2 are 0.48 and -0.18, respectively, so that the "average" values are $\bar{\rho}_1 = 0.70$ and $\bar{\rho}_2 = -0.31$. In our simulations, we accordingly consider paths of ρ_1 and ρ_2 that undergo a discrete shift in the middle of the sample, whose magnitude relative to this estimate is described by K, that is:

$$\rho_{1,t} = 0.70 + 0.48K\left(\mathbf{1}[t > T/2] - \tfrac{1}{2}\right) \quad \text{and} \quad \rho_{2,t} = -0.31 - 0.18K\left(\mathbf{1}[t > T/2] - \tfrac{1}{2}\right). \quad (17)$$

Under equation (17), the AR(2) process is stationary throughout when $0 \leq K \leq 2.5$. Regarding the second moments of the disturbances, we determine $E[v_t^2] = 0.55$, $E[\xi_t^2] = 2.11$, and $E[\xi_t v_t] = -0.01$ from the full-sample OLS estimates of the reduced form (15), and compute the implied values of $E[\varepsilon_t^2]$ and $E[\varepsilon_t \xi_t]$. The initial values s_0 and s_{-1} are set to 0, and $T = 160$.

The left panel of Figure 2 depicts the rejection probability of 5% nominal level t tests t_ϕ and t_κ on the stable structural parameters ϕ and κ as a function of K, based on the usual GMM

6. $\Delta\pi_t$ and s_t are constructed using series from the DRI-McGraw Hill database. The annual rate of quarterly inflation is defined as $\pi_t = 400 \times (\ln P_t - \ln P_{t-1})$ where the measure of P_t is the price index of non-financial business sector (LGDPB in DRI database). Unemployment gap is defined as $s_t = u_t - \bar{u}_t$, where u_t is the unemployment rate and $\bar{u}_t$ is the non-accelerating inflation rate of unemployment (NAIRU). The series u_t is obtained by converting a monthly series of unemployment for all workers (LHUR in DRI data set) to the quarterly basis. The NAIRU series is constructed as a cubic spline in time, following Staiger, Stock and Watson (1997a, 1997b).

formulas and a Newey-West estimator for $\hat{V}_T$ with truncation parameter 4, ignoring the instability in the data generating process. The right panel similarly shows the rejection probability of the usual and modified Nyblom statistics $N_{\phi\kappa}$ and $M_{\phi\kappa}$, defined in analogy to equations (13) and (14), that test the stability of ϕ and κ. All empirical rejection probabilities are based on asymptotic critical values, using 25,000 replications. The four tests underreject in the stable model with $K = 0$, presumably due to the relative weakness of the instruments, combined with weak endogeneity. For $K = 1$, a nominally 5% level Nyblom test of the stability of all four parameter has 57% power, and a Nyblom test that is constructed to detect instabilities in ρ_1 and ρ_2 has 68% power. These tests are also undersized in the stable model, with rejection probabilities of 1·8% and 2·4%, respectively. The considered instability (17) with $K = 1$ is thus not negligible in the sense of remaining undetected with high probability. Nevertheless, as predicted by Theorem 1, part (i), the null rejection probabilities of t_ϕ and t_κ remain reasonably close to the nominal level for not too large values of K. Similarly, the modified stability test $M_{\phi\kappa}$ has rejection probability much closer to the nominal level than the unmodified statistic $N_{\phi\kappa}$. The non-monotonicity of their rejection probabilities as a function of K seems to be related to the relative strengthening of the instruments for larger K.

4. CONCLUSIONS

This paper addresses the question of how to conduct inference on a stable subset of parameters in a GMM model with time-varying parameters. We find that under general conditions, conventional GMM inference on parameters that ignore the instability remains asymptotically valid, as long as the instability is of moderate magnitude in the sense of not being detectable with probability 1. Usual tests for instability of a subset of parameters are usually affected by instabilities elsewhere, and we suggest a class of modified tests that do not suffer from this feature.

In practice, it might not always be easy to decide which parameters are stable and which are not. While our modified tests may be used to shed some empirical light on the issue, under the asymptotics considered in this paper, it is not possible to determine the subset of stable parameters from the data with probability 1, even in the limit. In some instances, economic theory might be useful in making this choice, as in the Euler condition example considered above. But even when such additional information is considered unreliable or absent, the results of this paper still considerably broaden the applicability of standard asymptotic inference for many time series GMM models. When conducting inference on a parameter of interest, it is not necessary to assume that all nuisance parameters remain constant through time.

APPENDIX

The proofs of Theorem 1 and Lemmas 1–3 are based on the following lemma.

Lemma 4. *If* (i) $\psi : [0, 1] \mapsto \mathbb{R}^d$ *is a non-stochastic, bounded, and piece-wise continuous function with at most a finite number of discontinuities and left and right limits everywhere;* (ii) *the stochastic $d \times d$ matrices $\{w_{T,t}\}$ satisfy $T^{-1} \sum_{t=1}^{[\lambda T]} w_{T,t} \xrightarrow{p} \int_0^\lambda \vartheta(l)dl$ for all $0 \leq \lambda \leq 1$ and some non-stochastic Riemann-integrable function $\vartheta : [0, 1] \mapsto \mathbb{R}^{d \times d}$ for which $\sup_{0 \leq \lambda \leq 1} \|\vartheta(\lambda)\| < \infty$;* (iii) $T^{-1} \sum_{t=1}^{T} \|w_{T,t}\| = O_p(1)$ *and $\sup_{t \leq T} T^{-1} \|w_{T,t}\| \xrightarrow{p} 0$ and* (iv) *the stochastic $d \times d$ matrices $\{\tilde{w}_{T,t}\}$ satisfy $T^{-1} \sum_{t=1}^{T} \|\tilde{w}_{T,t} - w_{T,t}\| \xrightarrow{p} 0$, then for all $s \in [0, 1]$:*

$$T^{-1} \sum_{t=1}^{[sT]} \tilde{w}_{T,t} \psi(t/T) \xrightarrow{p} \int_0^s \vartheta(l)\psi(l)dl.$$

Furthermore, if (ii) *is strengthened to $\sup_{\lambda \in [0,1]} \|T^{-1} \sum_{t=1}^{[\lambda T]} w_{T,t} - \int_0^\lambda \vartheta(l)dl\| \xrightarrow{p} 0$, then $\sup_{s \in [0,1]} \|T^{-1} \sum_{t=1}^{[sT]} \tilde{w}_{T,t} \psi(t/T) - \int_0^s \vartheta(l)\psi(l)dl\| \xrightarrow{p} 0$.*

Proof.   We need to show that for all $\eta_1, \eta_2 > 0$, there exists T^* such that for all $T > T^*$, $P\big(\|T^{-1}\sum_{t=1}^{[sT]}\tilde{w}_{T,t}$ $\psi(t/T) - \int_0^s \vartheta(l)\psi(l)dl\| > \eta_1\big) < \eta_2$. Pick $\delta > 0$ small enough and T_1^* large enough such that $\delta\sup_{0\leq\lambda\leq 1}\|\vartheta(\lambda)\| < \eta_1/4$ and $P\big(\delta T^{-1}\sum_{t=1}^{T}\|w_{T,t}\| > \eta_1/4\big) < \eta_2/4$ for all $T > T_1^*$. Since continuity on a compact interval implies uniform continuity on that interval, ψ is continuous except at a finite number of points and ψ has left and right limits everywhere, the function ψ can be uniformly approximated by a sequence of step functions. Hence, there exists mutually disjoint intervals $\mathcal{I}_1,\dots,\mathcal{I}_N$, $N < \infty$, satisfying $\bigcup_i \mathcal{I}_i = [0,1]$ and bounded vectors $c_1,\dots,c_N$ such that $\varphi(\lambda) = \sum_{i=1}^N \mathbf{1}[\lambda \in \mathcal{I}_i]c_i$ and $\sup_{0\leq\lambda\leq 1}\|\psi(\lambda) - \varphi(\lambda)\| < \delta$. We have:

$$\left\|T^{-1}\sum_{t=1}^{[sT]}\tilde{w}_{T,t}\psi(t/T) - \int_0^s \vartheta(l)\psi(l)dl\right\| \leq \left\|T^{-1}\sum_{t=1}^{[sT]}(\tilde{w}_{T,t} - w_{T,t})\psi(t/T)\right\| + \left\|T^{-1}\sum_{t=1}^{[sT]}w_{T,t}(\psi(t/T) - \varphi(t/T))\right\|$$

$$+ \left\|T^{-1}\sum_{t=1}^{[sT]}w_{T,t}\varphi(t/T) - \int_0^s \vartheta(l)\varphi(l)dl\right\| + \left\|\int_0^s \vartheta(l)\varphi(l)dl - \int_0^s \vartheta(l)\psi(l)dl\right\|.$$

But:

$$\left\|\int_0^s \vartheta(l)\varphi(l)dl - \int_0^s \vartheta(l)\psi(l)dl\right\| \leq \delta\int_0^1 \|\vartheta(l)\|dl \leq \eta_1/4$$

$$\left\|T^{-1}\sum_{t=1}^{[sT]}w_{T,t}(\psi(t/T) - \varphi(t/T))\right\| < \delta T^{-1}\sum_{t=1}^{T}\|w_{T,t}\|$$

$$\left\|T^{-1}\sum_{t=1}^{[sT]}(\tilde{w}_{T,t} - w_{T,t})\psi(t/T)\right\| \leq \sup_{0\leq\lambda\leq 1}\|\psi(\lambda)\| \cdot T^{-1}\sum_{t=1}^{T}\|\tilde{w}_{T,t} - w_{T,t}\| \xrightarrow{p} 0,$$

so that the first result follows if we can show that $\left\|T^{-1}\sum_{t=1}^{[sT]}w_{T,t}\varphi(t/T) - \int_0^s \vartheta(l)\varphi(l)dl\right\| \xrightarrow{p} 0$. Now:

$$T^{-1}\sum_{t=1}^{[sT]}w_{T,t}\varphi(t/T) = T^{-1}\sum_{t=1}^{[sT]}w_{T,t}\sum_{i=1}^N \mathbf{1}[t/T \in \mathcal{I}_i]c_i$$

$$= \sum_{i=1}^N T^{-1}\left(\sum_{t\leq[sT],t/T\in\mathcal{I}_i}w_{T,t}\right)c_i,$$

and

$$\left\|\sum_{i=1}^N\left(T^{-1}\sum_{t\leq[sT],t/T\in\mathcal{I}_i}w_{T,t}\right)c_i - \sum_{i=1}^N\left(\int_{\mathcal{I}_i}\mathbf{1}[l\leq s]\vartheta(l)dl\right)c_i\right\|$$

$$\leq \sup_{i\leq N}\|c_i\| \cdot \sum_{i=1}^N\left\|T^{-1}\sum_{t\leq[sT],t/T\in\mathcal{I}_i}w_{T,t} - \int_{\mathcal{I}_i}\mathbf{1}[l\leq s]\vartheta(l)dl\right\|.$$

If the i-th interval is of the form $\mathcal{I}_i = (a_i, b_i]$, then $T^{-1}\sum_{t/T\in\mathcal{I}_i}w_{T,t} = T^{-1}\sum_{t=[a_iT]+1}^{[b_iT]}w_{T,t}$ and hence:

$$\left\|T^{-1}\sum_{t/T\in\mathcal{I}_i}w_{T,t} - \int_{\mathcal{I}_i}\vartheta(l)dl\right\| \leq \left\|T^{-1}\sum_{t=1}^{[b_iT]}w_{T,t} - \int_0^{b_i}\vartheta(l)dl\right\| + \left\|T^{-1}\sum_{t=1}^{[a_iT]}w_{T,t} - \int_0^{a_i}\vartheta(l)dl\right\| \xrightarrow{p} 0,$$

by assumption (ii). If the i-th interval is of the form $\mathcal{I}_i = [a_i, b_i)$, then:

$$\left\|T^{-1}\sum_{t/T\in\mathcal{I}_i}w_{T,t}\right\| \leq \left\|T^{-1}\sum_{t=[a_iT]+1}^{[b_iT]}w_{T,t}\right\| + T^{-1}\|w_{T,[a_iT]}\| + T^{-1}\|w_{T,[b_iT]}\|,$$

and $\left\| T^{-1} \sum_{t/T \in \mathcal{I}_i} w_{T,t} - \int_{\mathcal{I}_i} \vartheta(l) dl \right\| \xrightarrow{p} 0$ follows from the result just established and assumption (iii). The same arguments apply to the two other possible forms of the interval $\mathcal{I}_i$, and also to the interval $\mathcal{I}_i$ that contains s. Since N is fixed and finite, this implies:

$$T^{-1} \sum_{t=1}^{[sT]} w_{T,t} \varphi(t/T) \xrightarrow{p} \sum_{i=1}^{N} \left(\int_{\mathcal{I}_i} \mathbf{1}[l \leq s] \vartheta(l) dl \right) c_i = \int_0^s \vartheta(l) \varphi(l) dl.$$

For the second claim, proceed as above, and note that:

$$\sup_{s \in [0,1]} \sum_{i=1}^{N} \left\| T^{-1} \sum_{t \leq [sT], t/T \in \mathcal{I}_i} w_{T,t} - \int_{\mathcal{I}_i} \mathbf{1}[l \leq s] \vartheta(l) dl \right\| \leq 2N \sup_{\lambda \in [0,1]} \left\| T^{-1} \sum_{t=1}^{[\lambda T]} w_{T,t} - \int_0^{\lambda} \vartheta(l) dl \right\| \xrightarrow{p} 0. \quad \|$$

Proof of Theorem 1.   Since g is differentiable on Θ_0, and $\hat{\theta} \xrightarrow{p} \theta_0$, for large enough T and with probability converging to 1, the first-order condition of equation (6):

$$\left(T^{-1} \sum_{t=1}^{T} G_t(\hat{\theta}) \right)' Q_T T^{-1/2} \sum_{t=1}^{T} g_t(\hat{\theta}) = 0 = \hat{\Gamma}' Q_T T^{-1/2} \sum_{t=1}^{T} g_t(\hat{\theta}), \tag{A.1}$$

is satisfied. Also, since $\hat{\theta} \xrightarrow{p} \theta_0$ and $\|\theta_t - \theta_0\| \to 0$, for large enough T and with probability converging to 1, all line segments between $\hat{\theta}$ and θ_t are subsets of Θ_0. Hence, for large enough T, by a first-order Taylor expansion of $g_t(\hat{\theta})$ around $g_t(\theta_t)$ and summation over $t = 1, \ldots, [\lambda T]$ for $0 \leq \lambda \leq 1$:

$$T^{-1/2} \sum_{t=1}^{[\lambda T]} g_t(\hat{\theta}) = T^{-1/2} \sum_{t=1}^{[\lambda T]} g_t(\theta_t) + T^{-1/2} \sum_{t=1}^{[\lambda T]} \tilde{G}_t(\hat{\theta} - \theta_t)$$

$$= T^{-1/2} \sum_{t=1}^{[\lambda T]} g_t(\theta_t) + T^{-1/2} (\sum_{t=1}^{[\lambda T]} \tilde{G}_t)(\hat{\theta} - \theta_0) - T^{-1} \sum_{t=1}^{[\lambda T]} \tilde{G}_t f(t/T),$$

where the j-th row of $\tilde{G}_t$ is the j-th row of G_t evaluated at some $\tilde{\theta}_{t,j}$ that lies on the line segment between θ_t and $\hat{\theta}$.

Since $\hat{\theta} \xrightarrow{p} \theta_0$, there exists neighbourhoods $\mathcal{T}_T \subset \Theta_0$ of θ_0 converging to $\{\theta_0\}$ for which $P(\hat{\theta} \in \mathcal{T}_T) \to 1$. For T large enough to ensure that $\mathcal{T}_T \subset \Theta_0$, with probability converging to 1,

$$T^{-1} \sum_{t=1}^{[\lambda T]} \left\| \tilde{G}_t - G_t(\theta_0) \right\| \leq p T^{-1} \sum_{t=1}^{T} \sup_{\theta \in \mathcal{T}_T} \| G_t(\theta) - G_t(\theta_0) \| \xrightarrow{p} 0,$$

by Condition 1, part (vi), so that by Condition 1, part (vii), $T^{-1} \sum_{t=1}^{[\lambda T]} \tilde{G}_t \xrightarrow{p} \lambda \Gamma$ for all $0 \leq \lambda \leq 1$. Also, we can apply Lemma 4 to $T^{-1} \sum_{t=1}^{[\lambda T]} \tilde{G}_t f(t/T)$ and find $T^{-1} \sum_{t=1}^{[\lambda T]} \tilde{G}_t f(t/T) \xrightarrow{p} \Gamma \int_0^{\lambda} f(l) dl$ for all $0 \leq \lambda \leq 1$. From the first-order condition (A.1), $\left\| \hat{\Gamma} - \Gamma \right\| \xrightarrow{p} 0$ and $\| Q_T - Q_0 \| \xrightarrow{p} 0$ we find with these results that:

$$T^{1/2}(\hat{\theta} - \theta_0) = \int f(l) dl - (\Gamma' Q_0 \Gamma)^{-1} \Gamma' Q_0 T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t) + o_p(1). \tag{A.2}$$

The first result now follows from Condition 1, part (iii), and the continuous mapping theorem (CMT). Since equation (A.2) implies $\left\| \hat{\theta} - \theta_0 \right\| = O_p(T^{-1/2})$, we have for all $0 \leq \lambda \leq 1$:

$$T^{-1/2} \sum_{t=1}^{[\lambda T]} g_t(\hat{\theta}) = T^{-1/2} \sum_{t=1}^{[\lambda T]} g_t(\theta_t) + T^{1/2} \lambda \Gamma (\hat{\theta} - \theta_0) - \Gamma \int_0^{\lambda} f(l) dl + o_p(1). \tag{A.3}$$

Substituting equation (A.2) in equation (A.3) and rearranging yields:

$$T^{-1/2} \sum_{t=1}^{[\lambda T]} g_t(\hat{\theta}) = T^{-1/2} \sum_{t=1}^{[\lambda T]} g_t(\theta_t) - \lambda \Gamma (\Gamma' Q_0 \Gamma)^{-1} \Gamma' Q_0 T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t)$$

$$- \Gamma \left(\int_0^{\lambda} f(l) dl - \lambda \int_0^1 f(l) dl \right) + R_T^*(\lambda),$$

where $R_T^*(\lambda) = o_p(1)$ for all $0 \leq \lambda \leq 1$. The second result now follows from setting $\lambda = 1$. For the third result, notice that with a strengthening of the point-wise convergence in Condition 1, part (vii), to uniform convergence over λ, $\sup_{\lambda \in [0,1]} \left\| T^{-1} \sum_{t=1}^{[\lambda T]} \tilde{G}_t - \lambda \Gamma \right\| \xrightarrow{P} 0$ and $\sup_{\lambda \in [0,1]} \left\| T^{-1} \sum_{t=1}^{[\lambda T]} \tilde{G}_t f(t/T) - \Gamma \int_0^{\lambda} f(l) dl \right\| \xrightarrow{P} 0$ from the second claim in Lemma 4, so that $\sup_{\lambda \in [0,1]} \| R_T^*(\lambda) \| = o_p(1)$. The result then follows from $T^{-1/2} \sum_{t=1}^{[\cdot T]} g_t(\theta_t) \Rightarrow V^{1/2} W(\cdot)$ and the CMT. ∥

Proof of Lemma 1.   All the following computations are under the stable model with density $\prod_{t=1}^T f_{T,t}$ $(y_{T,t}, y_{T,t-1}, \ldots, y_{T,1}; \beta_0)$. The likelihood ratio statistic between the unstable and the stable models is $LR_T = \exp\left[\sum_{t=1}^T (l_t(\beta_t) - l_t(\beta_0))\right]$. Let $\mathcal{B}_T = \{\beta : \|\beta - \beta_0\| \leq T^{-1/2} \sup_{0 \leq \lambda \leq 1} \|B(\lambda)\|\}$. For T large enough to ensure that $\mathcal{B}_T \subset \mathcal{B}_0$, from an exact second-order Taylor expansion:

$$LR_T = \exp\left[\sum_{t=1}^T s_t(\beta_0)'(\beta_t - \beta_0) + \frac{1}{2}\sum_{t=1}^T (\beta_t - \beta_0)' h_t(\tilde{\beta}_t)(\beta_t - \beta_0)\right],$$

where $\tilde{\beta}_t$ lies on the line segment between β_0 and β_t. From Condition 2, part (iv),

$$T^{-1}\sum_{t=1}^T \left\| h_t(\tilde{\beta}_t) - h_t(\beta_0) \right\| \leq T^{-1}\sum_{t=1}^T \sup_{\beta \in \mathcal{B}_T} \|h_t(\beta) - h_t(\beta_0)\| \xrightarrow{P} 0.$$

Therefore,

$$\sum_{t=1}^T (\beta_t - \beta_0)' h_t(\tilde{\beta}_t)(\beta_t - \beta_0) = T^{-1}\mathrm{tr}\sum_{t=1}^T h_t(\tilde{\beta}_t) B(t/T) B(t/T)'$$

$$\xrightarrow{P} -\mathrm{tr}\int \Upsilon(l) B(l) B(l)' dl = -\int B(l)' \Upsilon(l) B(l) dl,$$

from a column-wise application of Lemma 4.

Let $q_t = s_t(\beta_0)' B(t/T)$. Then $\{q_t, \mathfrak{F}_t\}$ is a m.d. array, and:

$$T^{-1}\sum_{t=1}^T E[|q_t|^{2+\nu} \mid \mathfrak{F}_{t-1}] \leq \sup_{0 \leq \lambda \leq 1} \|B(\lambda)\|^{2+\nu} T^{-1}\sum_{t=1}^T E[\|s_t(\beta_0)\|^{2+\nu} \mid \mathfrak{F}_{t-1}],$$

which is $O_p(1)$ by Condition 2, part (iii). Also:

$$T^{-1}\sum_{t=1}^T E[q_t^2 \mid \mathfrak{F}_{t-1}] = T^{-1}\sum_{t=1}^T B(t/T)' E[s_t(\beta_0)s_t(\beta_0)' \mid \mathfrak{F}_{t-1}] B(t/T)$$

$$= \mathrm{tr}\, T^{-1}\sum_{t=1}^T E[s_t(\beta_0)s_t(\beta_0)' \mid \mathfrak{F}_{t-1}] B(t/T) B(t/T)'$$

$$\xrightarrow{P} \mathrm{tr}\int \Upsilon(l) B(l) B(l)' dl = \int B(l)' \Upsilon(l) B(l) dl,$$

where the convergence in probability stems from a column-wise application of Lemma 4. By corollary 3.1 in Hall and Heyde (1980), we hence have:

$$T^{-1/2}\sum_{t=1}^T q_t \Rightarrow \mathcal{N}(0, \omega^2),$$

where $\omega^2 = \int B(l)' \Upsilon(l) B(l) dl$. By the CMT, we conclude:

$$LR_T \Rightarrow \exp\left[\omega \mathcal{N}(0,1) - \frac{1}{2}\omega^2\right],$$

and contiguity follows after noting that $E \exp\left[\omega \mathcal{N}(0,1) - \frac{1}{2}\omega^2\right] = 1$ from LeCam's first lemma (van der Vaart, 1998, p. 88). ∥

Contiguity for stochastic parameter paths: Define $f_T(\{\beta_{T,t}\}_{t=1}^T) = \prod_{t=1}^T f_{T,t}(y_{T,t}, y_{T,t-1}, \ldots, y_{T,1}; \beta_{T,t})$, the density of $\{y_{T,t}\}_{t=1}^T$ with respect to the σ finite measure μ_T, let E_B stand for the integration over the probability measure of B and let A_T be the indicator function of a sequence of events with zero asymptotic probability in the

stable model, that is $\int A_T f_T(\{\beta_0\}_{t=1}^T) d\mu_T \to 0$. By (one equivalent) definition of contiguity (van der Vaart, 1998, p. 87), we need to show that A_T has asymptotic probability 0 also in the model with random parameter path $\{\beta_0 + T^{-1/2} B(t/T)\}_{t=1}^T$, that is $\int A_T E_B f_T(\{\beta_0 + T^{-1/2} B(t/T)\}_{t=1}^T) d\mu_T \to 0$. By Fubini's theorem, this is equivalent to $E_B \int A_T f_T(\{\beta_0 + T^{-1/2} B(t/T)\}_{t=1}^T) d\mu_T \to 0$, which follows from $\int A_T f_T(\{\beta_0 + T^{-1/2} b(t/T)\}_{t=1}^T) d\mu_T \to 0$ for almost all realizations $B = b$ by Lemma 1 and the dominated convergence theorem, since for all b, $0 \le \int A_T f_T(\{\beta_0 + T^{-1/2} b(t/T)\}_{t=1}^T) d\mu_T \le 1$.

Proof of Lemma 2.   We first prove $T^{-1/2} \sum_{t=1}^T g_t(\theta_t) \Rightarrow \mathcal{N}(0, V)$ in the unstable model by applying corollary 2.7 of McLeish (1974) to $\{v_g' g_t(\theta_t)\}_{t=1}^T$ for an arbitrary fixed $v_g' v_g = 1$, which yields the desired result by the Cramer-Wold device. Note that $T^{-1} \sum_{t=1}^T E[\|g_t(\theta_t)\|^{2+\varsigma} \mid \mathfrak{G}_{t-1}] = O_p(1)$ in the unstable model implies $T^{-1} \sum_{t=1}^T E[\|g_t(\theta_t)\|^2 \mathbf{1}[\|g_t(\theta_t)\| > T^{1/2} a] \mid \mathfrak{G}_{t-1}] \xrightarrow{p} 0$ for all $0 < a < \infty$ in the unstable model. To invoke corollary 2.7 of McLeish (1974), it thus remains to show that $T^{-1/2} \sup_{t \le T} \|g_t(\theta_t)\| \xrightarrow{p} 0$ and $\|T^{-1} \sum_{t=1}^T g_t(\theta_t) g_t(\theta_t)' - V\| \xrightarrow{p} 0$ in the unstable model. These convergences in probability follow from contiguity if we can show that they hold in the stable model.

The following computations hence concern the stable model. By an exact Taylor expansion:

$$g_t(\theta_t) = g_t(\theta_0) + \bar{G}_t(\theta_t - \theta_0),$$

where the j-th row of $\bar{G}_t$ is the j-th row of $G_t(\cdot)$ evaluated at some θ on the line segment between θ_0 and θ_t.

We compute:

$$T^{-1/2} \sup_{t \le T} \|g_t(\theta_t)\| \le T^{-1/2} \sup_{t \le T} \|g_t(\theta_0)\| + \sup_{t \le T} T^{-1} \|\bar{G}_t\| \sup_{0 \le \lambda \le 1} \|f(\lambda)\|.$$

But $T^{-1/2} \sup_{t \le T} \|g_t(\theta_0)\| \xrightarrow{p} 0$ by assumption, and with $\Theta_T = \{\theta : \|\theta - \theta_0\| \le T^{-1/2} \sup_{0 \le \lambda \le 1} \|f(\lambda)\|\}$,

$$T^{-1} \sup_{t \le T} \|\bar{G}_t\| \le p T^{-1} \sup_{t \le T} \sup_{\theta \in \Theta_T} \|G_t(\theta) - G_t(\theta_0) + G_t(\theta_0)\|$$

$$\le p T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_T} \|G_t(\theta) - G_t(\theta_0)\| + p T^{-1} \sup_{t \le T} \|G_t(\theta_0)\|.$$

The second term is $o_p(1)$ by assumption, and the first term is $o_p(1)$ by Condition 1, part (vi). Also:

$$T^{-1} \sum_{t=1}^T g_t(\theta_t) g_t(\theta_t)' = T^{-1} \sum_{t=1}^T g_t(\theta_0) g_t(\theta_0)' + T^{-1} \sum_{t=1}^T g_t(\theta_0)(\theta_t - \theta_0)' \bar{G}_t'$$

$$+ T^{-1} \sum_{t=1}^T \bar{G}_t(\theta_t - \theta_0) g_t(\theta_t)' + T^{-1} \sum_{t=1}^T \bar{G}_t(\theta_t - \theta_0)(\theta_t - \theta_0)' \bar{G}_t',$$

where

$$T^{-1} \sum_{t=1}^T \|\bar{G}_t(\theta_t - \theta_0) g_t(\theta_t)'\| \le (\sup_{0 \le \lambda \le 1} \|f(\lambda)\|) T^{-1} \sum_{t=1}^T \|\bar{G}_t\| \cdot \|T^{-1/2} g_t(\theta_t)\|$$

$$\le (\sup_{0 \le \lambda \le 1} \|f(\lambda)\|)(T^{-1/2} \sup_{t \le T} \|g_t(\theta_t)\|) T^{-1} \sum_{t=1}^T \|\bar{G}_t\| \xrightarrow{p} 0,$$

since, as shown above, $T^{-1/2} \sup_{t \le T} \|g_t(\theta_t)\| \xrightarrow{p} 0$ and:

$$T^{-1} \sum_{t=1}^T \|\bar{G}_t\| \le p T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_T} \|G_t(\theta) - G_t(\theta_0) + G_t(\theta_0)\|$$

$$\le p T^{-1} \sum_{t=1}^T \sup_{\theta \in \Theta_T} \|G_t(\theta) - G_t(\theta_0)\| + p T^{-1} \sum_{t=1}^T \|G_t(\theta_0)\|,$$

which is $O_p(1)$ by Condition 1, part (vi). Finally:

$$T^{-1} \sum_{t=1}^T \|\bar{G}_t(\theta_t - \theta_0)(\theta_t - \theta_0)' \bar{G}_t'\| \le (\sup_{0 \le \lambda \le 1} \|f(\lambda)\|)^2 T^{-2} \sum_{t=1}^T \|\bar{G}_t\|^2$$

$$\le (\sup_{0 \le \lambda \le 1} \|f(\lambda)\|)^2 (T^{-1} \sup_{t \le T} \|\bar{G}_t\|) T^{-1} \sum_{t=1}^T \|\bar{G}_t\| \xrightarrow{p} 0.$$

For the second claim of the lemma, proceed as above to show that also $\left\| T^{-1} \sum_{t=1}^{[\lambda T]} g_t(\theta_t) g_t(\theta_t)' - \lambda V \right\| \xrightarrow{P} 0$ for each $0 \le \lambda \le 1$ in the stable model, so that the same condition holds under the unstable model by contiguity. The result now follows from theorem 3.6 in McLeish (1974) and the functional Cramer-Wold device (*cf.* Davidson, 1994, theorem 29.16). ‖

Proof of Lemma 3.   As in the proof of Lemma 1, all calculations are made under the stable model. From a first-order exact Taylor expansion:

$$T^{-1/2} \sum_{t=1}^{T} s_t(\beta_t) = T^{-1/2} \sum_{t=1}^{T} s_t(\beta_0) + T^{-1} \sum_{t=1}^{T} \tilde{h}_t B(t/T),$$

where the j-th row of $\tilde{h}_t$ is equal to the j-th row of $h_t(\cdot)$ evaluated at some $\tilde{\beta}_{t,j}$ on the line segment between β_0 and β_t, so that by the same arguments used in the proofs of Lemmas 1 and 2 above,

$$\left\| T^{-1/2} \sum_{t=1}^{T} s_t(\beta_t) - T^{-1/2} \sum_{t=1}^{T} s_t(\beta_0) + \int \Upsilon(l) B(l) dl \right\| \xrightarrow{P} 0.$$

Let the scalar v_0 and the $k \times 1$ vector v_1 be such that $v = (v_0, v_1')'$ satisfies $v'v = 1$. With $z_t = v_0 B(t/T)' s_t(\beta_0) + v_1' s_t(\beta_0)$, $\{z_t, \mathfrak{F}_t\}$ is a m.d. array with conditional variance:

$$E[z_t^2 \mid \mathfrak{F}_{t-1}] = (v_0 B(t/T) + v_1)' E[s_t(\beta_0) s_t(\beta_0)' \mid \mathfrak{F}_{t-1}] (v_0 B(t/T) + v_1).$$

Following the reasoning in the proof of Lemma 1 shows that corollary 3.1 of Hall and Heyde (1980) is applicable and we find:

$$T^{-1/2} \sum_{t=1}^{T} z_t \Rightarrow \mathcal{N}\left(0, \int (v_0 B(l) + v_1)' \Upsilon(l) (v_0 B(l) + v_1) dl\right).$$

Applying the Cramer-Wold device and the CMT, we therefore obtain:

$$\left(\ln \mathrm{LR}_T, T^{-1/2} \sum_{t=1}^{T} s_t(\beta_t)' \right)' \Rightarrow \mathcal{N}\left(\begin{pmatrix} -\frac{1}{2}\omega^2 \\ -\int \Upsilon(l) B(l) dl \end{pmatrix}, \begin{pmatrix} \omega^2 & \int B(l)' \Upsilon(l) dl \\ \int \Upsilon(l) B(l) dl & \int \Upsilon(l) dl \end{pmatrix} \right).$$

But by LeCam's third lemma (*cf.* van der Vaart, 1998, p. 90), this implies that under the unstable model,

$$T^{-1/2} \sum_{t=1}^{T} s_t(\beta_t) \Rightarrow \mathcal{N}\left(0, \int \Upsilon(l) dl\right),$$

and the result follows.   ‖

REFERENCES

ANDREWS, D. W. K. (1987), "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers", *Econometrica*, **55**, 1465–1471.
ANDREWS, D. W. K. (1991), "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, **59**, 817–858.
ANDREWS, D. W. K. (1992), "Generic Uniform Convergence", *Econometric Theory*, **8**, 241–257.
ANDREWS, D. W. K. (1993), "Tests for Parameter Instability and Structural Change with Unknown Change Point", *Econometrica*, **61**, 821–856.
ANDREWS, D. W. K. (2003), "End-of-Sample Instability Tests", *Econometrica*, **71**, 1661–1694.
ANDREWS, D. W. K. and PLOBERGER, W. (1994), "Optimal Tests When a Nuisance Parameter is Present Only under the Alternative", *Econometrica*, **62**, 1383–1414.
BAI, J. (1994), "Least Squares Estimation of a Shift in Linear Processes", *Journal of Time Series Analysis*, **15**, 453–470.
BAI, J. and PERRON, P. (1998), "Estimating and Testing Linear Models with Multiple Structural Changes", *Econometrica*, **66**, 47–78.

BLANCHARD, O. J. and GALI, J. (2007), "Real Wage Rigidities and the New Keynesian Model", *Journal of Money, Credit and Banking*, **39**, 35–65.

BOIVIN, J. (1999), "Revisiting the Evidence on the Stability of Monetary VARs" (Working Paper, Columbia University).

COGLEY, T. and SARGENT, T. J. (2005), "Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII US", *Review of Economic Dynamics*, **8**, 262–302.

DAVIDSON, J. (1994), *Stochastic Limit Theory* (New York: Oxford University Press).

ELLIOTT, G. and MÜLLER, U. K. (2006), "Efficient Tests for General Persistent Time Variation in Regression Coefficients", *Review of Economic Studies*, **73**, 907–940.

ELLIOTT, G. and MÜLLER, U. K. (2007), "Confidence Sets for the Date of a Single Break in Linear Time Series Regressions", *Journal of Econometrics*, **141**, 1196–1218.

GALLANT, A. R. and WHITE, H. (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models* (Oxford: Basil Blackwell).

GHYSELS, E. (1998), "On Stable Factor Structures in the Pricing of Risk: Do Time-Varying Betas Help or Hurt?", *Journal of Finance*, **53**, 549–573.

GHYSELS, E., GUAY, A. and HALL, A. (1997), "Predictive Tests for Structural Change with Unknown Breakpoint", *Journal of Econometrics*, **82**, 209–233.

GHYSELS, E. and HALL, A. (1990), "Are Consumption-Based Intertemporal Capital Asset Pricing Models Structural?", *Journal of Econometrics*, **45**, 121–139.

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.

HALL, P. and HEYDE, C. C. (1980), *Martingale Limit Theory and its Applications* (New York: Academic Press).

HANSEN, B. E. (1990), "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models" (Working Paper, Rochester University).

HANSEN, B. E. (2000), "Testing for Structural Change in Conditional Models", *Journal of Econometrics*, **97**, 93–115.

HANSEN, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators", *Econometrica*, **50**, 1029–1054.

KREPS, D. (1998), "Anticipated Utility and Dynamic Choice", in D. P. Jacobs, E. Kalai and M. I. Kamien (eds.) *Frontiers of Research in Economic Theory* (Cambridge, UK: Cambridge University Press) 242–274.

LI, H. (2008), "Estimation and Testing of Euler Equation Models with Time-Varying Reduced-Form Coefficients", *Journal of Econometrics*, **142**, 428–448.

LINDE, J. (2001), "Testing for the Lucas-Critique: A Quantitative Investigation", *American Economic Review*, **91**, 986–1005.

LUCAS, R. E. (1976), "Econometric Policy Evaluation: A Critique", *Carnegie Rochester Conference Series on Public Policy*, **1**, 19–46.

MCLEISH, D. L. (1974), "Dependent Central Limit Theorems and Invariance Principles", *The Annals of Probability*, **2**, 620–628.

MÜLLER, U. K. (2007), "An Alternative Sense of Asymptotic Efficiency" (Working Paper, Princeton University).

MÜLLER, U. K. and PETALAS, P. (2007), "Efficient Estimation of the Parameter Path in Unstable Time Series Models" (Working Paper, Princeton University).

NEWEY, W. K. (1985), "Generalized Method of Moments Specification Testing", *Journal of Econometrics*, **29**, 229–256.

NEWEY, W. K. and WEST, K. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, **55**, 703–708.

NYBLOM, J. (1989), "Testing for the Constancy of Parameters Over Time", *Journal of the American Statistical Association*, **84**, 223–230.

PHILLIPS, P. C. B. and PLOBERGER, W. (1996), "An Asymptotic Theory of Bayesian Inference for Time Series", *Econometrica*, **64**, 381–412.

PLOBERGER, W., KRÄMER, W. and KONTRUS, K. (1989), "A New Test for Structural Stability in the Linear Regression Model", *Journal of Econometrics*, **40**, 307–318.

PRIMICERI, G. E. (2005), "Time Varying Structural Vector Autoregressions and Monetary Policy", *The Review of Economic Studies*, **72**, 821–852.

SOWELL, F. (1996), "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework", *Econometrica*, **64**, 1085–1107.

STAIGER, D., STOCK, J. H. and WATSON, M. W. (1997a), "How Precise are Estimates of the Natural Rate of Unemployment", in C. Romer and D. Romer (eds.) *Reducing Inflation: Motivation and Strategy* (Chicago: University of Chicago Press) 195–242.

STAIGER, D., STOCK, J. H. and WATSON, M. W. (1997b), "The NAIRU, Unemployment, and Monetary Policy", *Journal of Economic Perspectives*, **11**, 33–51.

STOCK, J. H. and WATSON, M. W. (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations", *Journal of Business and Economic Statistics*, **14**, 11–30.

STOCK, J. H. and WATSON, M. W. (1998), "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model", *Journal of the American Statistical Association*, **93**, 349–358.

VAN DER VAART, A. W. (1998), *Asymptotic Statistics* (Cambridge, UK: Cambridge University Press).