Valid Inference in Partially Unstable GMM Models

Li, Hong and Mueller, Ulrich

Princeton University

August 2006

Valid Inference in Partially Unstable GMM Models*

Hong Li
Department of Economics
Brandeis University
Waltham, MA 02454, USA
hli@brandeis.edu

Ulrich K. Müller†
Department of Economics
Princeton University
Princeton, NJ 08544, USA
umueller@princeton.edu

Abstract

The paper considers time series GMM models where a subset of the parameters are time varying. The magnitude of the time variation in the unstable parameters is such that efficient tests detect the instability with (possibly high) probability smaller than one, even in the limit. We show that for many forms of the instability and a large class of GMM models, standard GMM inference on the subset of stable parameters, ignoring the partial instability, remains asymptotically valid.

JEL Classification: C32

Keywords: Structural Breaks, Parameter Stability Test, Contiguity, Euler Condition, New Keynesian Phillips Curve

*We thank Graham Elliott and Mark Watson, as well as seminar participants at Houston University, U.C. Davis, Brandeis University, the Federal Reserve Board, the NY Federal Reserve Bank, the Atlanta Federal Reserve Bank, Iowa State University, the Econometric Society World Congress 2005 and the University of Pennsylvania for very helpful discussions. Müller gratefully acknowledges financial support by NSF grant SES-0518036.

†Corresponding author. Tel: +1 609 258 4026; fax: +1 609 258 6419.

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 behavioral changes of rational economic agents, which results 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 macroeconomic 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 (2005) 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 one 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.

Here we analyze models where only a subset of parameters are unstable, and focus on instabilities that are small in the sense that reasonable tests detect them with (possibly large) probability smaller than one in the limit. 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 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. The key assumption is that sample averages of the derivative of the moment condition are approximately the same for all parts of the sample. This holds for most 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 stable GMM model are Euler moment conditions of optimizing agents under a time varying policy environment. Rational economic agents adapt their optimal behavior 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. One application of this paper's result is how to conduct inference about this subset of stable parameters; see Li (2004) for an application to an investment model and Section 4 below for a stylized New Keynesian Phillips Curve example.

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 instabilities in other parts of the model.

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. We emphasize that this does not mean that our results only apply to economically insignificant instabilities. 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 macroeconomic data series documented in Stock and Watson (1996) are often only borderline significant. In such instances, accurate approximations are generated by a modelling strategy in which there is only limited information about the instability asymptotically, as in the $T^{-1/2}$ neighborhood. And indeed, in our Monte Carlo simulations we find that our asymptotic results provide accurate approximations for instabilities that are large by empirical standards.

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.

On a technical level, the analysis of time series models with time varying parameters faces the difficulty that these models tend to generate nonstationary 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 (2005), 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 for the asymptotic analysis of unstable time series models.

The next section introduces the model and discusses a set of high-level conditions on the partially unstable GMM model. These high-level conditions on the unstable model are then justified by appropriate assumptions about the properties of the corresponding stable GMM model. Section 3 contains the main result, and discusses its implications for econometric practice. In Section 4 we consider the small sample relevance of the main result in a Monte Carlo study. Section 5 concludes. Proofs are collected in an Appendix.

2 Model and Assumptions

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 is given by a triangular array of 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

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

For notational convenience, we will drop the dependence of $y_{T,t}$ and $\theta_{T,t}$ on T if no confusion arises. Also, let $g_t(\theta)$ be $g(y_t, \theta_t)$. 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 zero and one, if not indicated otherwise.

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

$$\left[T^{-1}\sum_{t=1}^{T} g_t(\hat{\theta})\right]' Q_T \left[T^{-1}\sum_{t=1}^{T} g_t(\hat{\theta})\right] = \inf_{\theta\in\Theta} \left[T^{-1}\sum_{t=1}^{T} g_t(\theta)\right]' Q_T \left[T^{-1}\sum_{t=1}^{T} g_t(\theta)\right], \tag{2}$$

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*

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

(ii) In some neighborhood 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 neighborhood* Θ_T *of* θ_0 *contained in* Θ_0, *i.e.* $\Theta_T = \{\theta : ||\theta - \theta_0|| < c_T\} \subset \Theta_0$ *for some sequence of real numbers* $c_T \to 0$, $T^{-1}\sum_{t=1}^T \sup_{\theta\in\Theta_T} ||G_t(\theta) - G_t(\theta_0)|| \xrightarrow{p} 0$.

(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 neighborhood in which efficient tests of parameter stability have nontrivial 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 nonstochastic, in contrast to, say, Stock and Watson (1998) and Elliott and Müller (2005). But under an alternative assumption of stochastic parameter paths, the following results continue to hold as long as Condition 1 holds for almost all realizations of the path. Almost all realizations of a Wiener process on the unit interval, for instance, are bounded and continuous, and hence may serve as functions f as specified in part (i).

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 (1), 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 $\{g_t(\theta_t)\}_{t=1}^T$ need to checked in the unstable model, a complication to which we return below.

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, i.e. if f was equal to zero. 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 nonstochastic 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 nonzero f, these convergences in probability are less standard, and we provide a suitable argument below.

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 nonstationary 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 (vii) of Condition 1, even after scale normalizations that ensure $T^{-1}\sum_{t=1}^{T} G_t(\theta_0) = O_p(1)$.

As noted above, assumptions in parts (iii)–(vii) are fairly standard for stable GMM models. The analysis of unstable models is complicated by the fact that parameter instability typically leads to nonstationary 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, Krämer, and Kontrus (1989) only consider regression models with strictly exogenous regressors. Sowell (1996) assumes that both the stable and unstable model 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 (an analysis of the behavior of long-run variance estimators under neglected non-local parameter shifts is provided by Hall, Inoue, and Peixe (2003)).

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 (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 unstable model 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 (iv)–(vii) to be true in the unstable GMM model whenever they hold in the corresponding stable GMM model (i.e. when $f = 0$). Formally, a sequence of probability models is called *contiguous* to another sequence of probability models defined on the same probability space whenever all $o_p(1)$ statistics under the latter remain $o_p(1)$ under the former—see van der Vaart (1998), Chapter 6 and Pollard (2001) for further discussion.

To make the above heuristic reasoning rigorous, we need to impose some regularity conditions on the generating process of the data $\{y_{T,t}\}_{t=1}^T$. Assume that the difference between the density of the stable and unstable model can be described by the evolution of the $k \times 1$ parameter β, $k \geq p$, 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 $\prod_{t=1}^s f_{T,t}(y_{T,t}, y_{T,t-1}, \cdots, y_{T,1}; \beta_{T,t})$ when β takes on the value $\beta_{T,t}$ at date t. With $k > p$, this allows the instability in the likelihood to go beyond the instability in the GMM parameter θ. Denote by $l_{T,t}(\beta) = \ln f_{T,t}(y_{T,t}, y_{T,t-1}, \cdots, y_{T,1}; \beta)$ 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 simplicity. Also, we refer to the model with density $\prod_{t=1}^T f_{T,t}(y_{T,t}, y_{T,t-1}, \cdots, y_{T,1}; \beta_0)$ 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 piecewise continuous vector function $B : [0,1] \mapsto \mathbb{R}^k$ with at most a finite number of discontinuities.*

(ii) In some neighborhood $\mathcal{B}_0$ of β_0, $l_t(\beta)$ is twice differentiable a.s. with respect to β for

$t = 1, \cdots, T$.

Furthermore, in the stable model,

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

(iv) $T^{-1}\sum_{t=1}^{T}||h_t(\beta_0)|| = O_p(1)$, $T^{-1}\sup_{t\leq T}||h_t(\beta_0)|| \xrightarrow{p} 0$ *and for any decreasing neighborhood* $\mathcal{B}_T$ *of* β_0 *contained in* $\mathcal{B}_0$, $T^{-1}\sum_{t=1}^{T}\sup_{\beta\in\mathcal{B}_T}||h_t(\beta) - h_t(\beta_0)|| \xrightarrow{p} 0$.

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

Part (i) makes the same assumption on the form of the instability in β as Condition 1 (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 $T^{-1}\sum_{t=1}^{T} E[s_t(\beta_0)s_t(\beta_0)'|\mathfrak{F}_{t-1}]$. 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 Conditions 1 (iv)–(vii) and 2, then Condition 1 (iv)–(vii) also holds under the unstable model.*

Lemma 1 formally states the possibility of obtaining Condition 1 (iv)–(vii) by making assumptions only on the stable GMM model. As argued above, Condition 1 (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 set-up, Condition 2 plays the role of a regularity condition, akin to more familiar mixing or moment conditions.

Much applied work, such as Stock and Watson (1996), Cogley and Sargent (2005) and Primiceri (2005), proceed under an alternative assumption of stochastic parameter paths. Our results are still applicable in this scenario, as long as Condition 2 holds for almost all realizations of B, that is the stochastic parameter path is independent of the model disturbances in the corresponding stable model. 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. See the appendix for a detailed argument for contiguity of the unstable model to the corresponding stable model with stochastic parameter paths.

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 (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 $\epsilon > 0$ such that $T^{-1}\sum_{t=1}^T E[||g_t(\theta_t)||^{2+\epsilon}|\mathfrak{G}_{t-1}] = O_p(1)$, 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 $T^{-1}\sum_{t=1}^T g_t(\theta_0)g_t(\theta_0)' \xrightarrow{p} V$, 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 $T^{-1}\sum_{t=1}^{[\lambda T]} g_t(\theta_t)g_t(\theta_t)' \xrightarrow{p} \lambda V$ for all $0 \leq \lambda \leq 1$, 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 $\sup_t E[||\varepsilon_t||^{2+\epsilon}|\mathfrak{G}_{t-1}] \leq \bar{M}_\varepsilon$ a.s under the unstable model. Then $T^{-1}\sum_{t=1}^T E[||g_t(\theta_t)||^{2+\epsilon}|\mathfrak{G}_{t-1}] \leq \bar{M}_\varepsilon T^{-1}\sum_{t=1}^T ||x_{t-1}||^{2+\epsilon}$ a.s. in the unstable model, and it suffices to show that $T^{-1}\sum_{t=1}^T ||x_{t-1}||^{2+\epsilon} = O_p(1)$ 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 are 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* $||T^{-1/2}\sum_{t=1}^T g_t(\theta_0) - T^{-1/2}F'\sum_{t=1}^T s_t(\beta_0)|| \xrightarrow{p} 0$ *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 (iv)–(vii) also holds under the unstable model. In general, Condition 1 (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 (iii) holds in the stable model. The following asymptotic results thus hold 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 Asymptotic Results

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,* $\sup_{\lambda\in[0,1]} ||T^{-1}\sum_{t=1}^{[\lambda T]} G_t(\theta_0) - \lambda\Gamma|| \xrightarrow{p} 0$ *and* $T^{-1/2}\sum_{t=1}^{[\cdot T]} g_t(\theta_t) \Rightarrow V^{1/2}W(\cdot)$*, then*

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

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. Wald statistics involving only stable parameters are asymptotically chi-squared under the null hypothesis, and have the same noncentrality parameter under local alternatives as the corresponding fully stable model. It is immediate from Condition 1 (i) and (iv) that the GMM estimator $\hat{\theta}$ is consistent for the average parameter value, $||\hat{\theta} - T^{-1}\sum_{t=1}^T \theta_t|| \xrightarrow{p} 0$. 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 (2)

$$\begin{aligned} 0 &= \hat{\Gamma}' Q_T T^{-1/2}\sum_{t=1}^T g_t(\hat{\theta}) \\ &= \hat{\Gamma}' Q_T T^{-1/2}\sum_{t=1}^T g_t(\theta_t) + \hat{\Gamma}' Q_T (T^{-1}\sum_{t=1}^T \tilde{G}_t) T^{1/2}(\hat{\theta} - \theta_0) - \hat{\Gamma}' Q_T T^{-1}\sum_{t=1}^T \tilde{G}_t T^{1/2}(\theta_t - \theta_0) \end{aligned} \tag{3}$$

where the jth row of $\tilde{G}_t$ is the jth 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 (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 (i) now follows from rearranging (3) and taking limits, revealing the relevance of this result for conducting asymptotically valid inference in partially stable models.

To develop a better intuition, consider the linear regression model

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

where X_t and Z_t are two (possibly correlated) scalar random variables. By standard OLS algebra (the Frisch-Waugh Theorem), the standard t-statistic on θ_2 is numerically identical to the t-statistic on $\tilde{\theta}_2$ in the model

$$Y_t = X_t \theta_{1,t} + \tilde{Z}_t \tilde{\theta}_2 + \varepsilon_t = X_t \theta_{1,0} + \tilde{Z}_t \tilde{\theta}_2 + X_t(\theta_{1,t} - \theta_{1,0}) + \varepsilon_t$$

where $\tilde{Z}_t$ is a (nonzero) linear combination of X_t and Z_t which is uncorrelated with X_t. If (X_t, Z_t) is stationary 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 $\tilde{Z}_t$. Inference on $\tilde{\theta}_2$ (and hence θ_2) thus 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 $\tilde{\theta}_2$ (and hence θ_2).

As a consequence of part (ii) of Theorem 1, Hansen's (1982) overidentification test 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.

Part (iii) of Theorem 1 requires the strengthening of Condition 1 (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 (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 analyzed in Nyblom (1989), Sowell (1996) and Elliott and Müller (2005), 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 stable 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 efficient tests of

$$H_0 : \theta_t \text{ is constant in } t \quad \text{ against } \quad H_1 : \theta_t \text{ depends on } t \tag{4}$$

in the class of tests that are continuous functions of $T^{-1/2}\sum_{t=1}^{[\cdot T]} g_t(\hat{\theta})$. Specifically, Sowell's Corollary 2 shows that tests of (4) that maximize power against alternatives where only the first m_0 elements of θ are time varying are functions of

$$\begin{aligned}(C'\hat{\Sigma}_\theta^{-1}C)^{-1/2}C'\hat{\Gamma}'\hat{V}_T^{-1}T^{-1/2}\sum_{t=1}^{[\cdot T]} g_t(\hat{\theta}) &\Rightarrow W_{m_0}(\cdot) - \cdot W_{m_0}(1) \\ &+ (C'\Sigma_\theta^{-1}C)^{-1/2}C'\Sigma_\theta^{-1}\left(\int_0^{\cdot} f(l)dl - \cdot \int_0^1 f(l)dl\right)\end{aligned} \tag{5}$$

where $\Sigma_\theta = (\Gamma' V^{-1}\Gamma)^{-1}$ and W_{m_0} is a $m_0 \times 1$ standard Wiener process, so that $W_{m_0}(\lambda) - \lambda W_{m_0}(1)$ is a $m_0 \times 1$ Brownian Bridge. In general, as long as Σ_θ is not block diagonal, the asymptotic distribution (5) 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 the more specific hypothesis

$$H_0 : C'\theta_t \text{ is constant in } t \quad \text{ against } \quad H_1 : C'\theta_t \text{ depends on } t \tag{6}$$

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

$$\begin{aligned}(C'\hat{\Sigma}_\theta C)^{-1/2}C'\hat{\Sigma}_\theta\hat{\Gamma}'\hat{V}_T^{-1}T^{-1/2}\sum_{t=1}^{[\cdot T]} g_t(\hat{\theta}) \quad &\Rightarrow \quad W_{m_0}(\cdot) - \cdot W_{m_0}(1) \\ &+ (C'\Sigma_\theta C)^{-1/2}C'\left(\int_0^{\cdot} f(l)dl - \cdot \int_0^1 f(l)dl\right).\end{aligned} \tag{7}$$

The stochastic part of the asymptotic distribution is again a standard Brownian Bridge of dimension m_0. When one applies the same type of test statistic to (5) and (7), 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 left-hand side of (5) and (7), one obtains the same asymptotic distribution when all parameters are stable, and thus the same critical value. But in contrast to (5), under the restricted null hypothesis (6) of stability of the first m_0 elements of θ, the second summand in (7) 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 (7) will not be affected by instabilities in the non-tested parameters. If a stability test based on a functional of (7) 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 (7) typically have lower power than tests based on (5): 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 (5) 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 (7). For tests that seek to detect potential instabilities in all parameters, i.e. $C = I_m$, (7) reduces to (5).

In summary, for a partially stable 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 (7), whose asymptotic rejection probabilities are a function of the stability of the parameters under consideration only.

4 Monte Carlo Results

The results of the last section show that usual GMM inference about a stable subset of parameters in a locally unstable model remains asymptotically valid if 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 considers the linear regression example considered above, augmented for a constant term

$$Y_t = X_t\theta_{1,t} + Z_t\theta_2 + \theta_3 + \varepsilon_t,\ t = 1, \cdots, T \tag{8}$$

where $\varepsilon_t \sim i.i.d.\ \mathcal{N}(0,1)$ and $(X_t, Z_t)'$ is a zero-mean stationary Gaussian VAR(1) 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 (8). This is an exactly identified GMM problem where $\hat{\Gamma} = T^{-1}\sum_{t=1}^{T} R_tR_t'$ and, for heteroskedasticity robust inference, $\hat{V}_T = T^{-1}\sum_{t=1}^{T} R_tR_t'\hat{\varepsilon}_t^2$.

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 (5) is then given by[1]

$$\mathrm{N} = T^{-1}\sum_{s=1}^{T}\left(C'\hat{\Gamma}'\hat{V}_T^{-1}\sum_{t=1}^{s}R_t\hat{\varepsilon}_t\right)'\left(C'\hat{\Sigma}_\theta^{-1}C\right)^{-1}\left(C'\hat{\Gamma}'\hat{V}_T^{-1}\sum_{t=1}^{s}R_t\hat{\varepsilon}_t\right) \tag{9}$$

and the modified Nyblom statistic based on (7) is

$$\mathrm{M} = T^{-1}\sum_{s=1}^{T}\left(C'\hat{\Sigma}_\theta\hat{\Gamma}'\hat{V}_T^{-1}\sum_{t=1}^{s}R_t\hat{\varepsilon}_t\right)'\left(C'\hat{\Sigma}_\theta C\right)^{-1}\left(C'\hat{\Sigma}_\theta\hat{\Gamma}'\hat{V}_T^{-1}\sum_{t=1}^{s}R_t\hat{\varepsilon}_t\right). \tag{10}$$

By Theorem 1 (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 the tables) correspond to $h = 5$ and $h = 8$ in the single break case and the random walk case, respectively; large instabilities (denoted as 'lg' in the tables) correspond to $h = 10$ and

[1] This version of the heteroskedasticity robust Nyblom (1989) statistic differs from what is suggested in Hansen (1990) and often employed in practice, that is $T^{-1}\sum_{s=1}^{T}(C'\sum_{t=1}^{s}R_t\hat{\varepsilon}_t)'\left(T^{-1}\sum_{t=1}^{T}C'R_tR_t'C\hat{\varepsilon}_t^2\right)^{-1}(C'\sum_{t=1}^{s}R_t\hat{\varepsilon}_t)$. The optimality result of Sowell (1996) discussed above implies that in the presence of heteroskedasticity, (9) is the more powerful statistic, at least asymptotically.

Table 1: Small Sample Rejection Probabilities of 5% Nominal Tests in Percent, $\rho_{XZ} = 0.5$

	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
							$r = 0$							
0	6.5	6.3	4.6	4.9	5.0	4.9	4.9	6.4	6.3	4.6	5.0	5.0	5.0	4.7
sm	6.2	6.3	34.7	53.8	15.1	43.1	4.5	5.9	6.4	34.0	44.1	14.3	38.5	4.6
lg	5.5	6.3	89.4	97.2	30.8	92.4	3.4	5.2	6.5	58.6	68.7	22.0	63.9	4.2
							$r = 0.5$							
0	6.4	6.4	4.3	4.7	4.6	4.6	4.7	6.4	6.2	4.3	4.6	4.6	4.6	4.6
sm	6.8	6.8	32.1	50.4	14.4	39.2	4.4	6.9	7.0	33.2	42.5	14.2	36.6	4.7
lg	7.6	8.0	86.9	96.0	30.6	89.5	3.6	7.6	8.3	59.4	67.2	22.6	62.2	5.1
							$r = 0.95$							
0	6.2	6.0	2.1	2.0	1.9	2.6	2.5	6.4	6.4	2.2	2.2	2.0	2.9	2.7
sm	13.0	12.3	13.6	20.6	8.3	12.8	3.7	15.3	14.2	19.5	23.1	10.6	16.6	6.2
lg	26.3	24.2	48.0	60.5	23.3	41.7	8.0	26.5	25.5	46.0	47.9	23.0	38.0	14.8

$h = 16$. We set the sample size $T = 100$ and, as a benchmark, $\rho_{XZ} = 0.5$. 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 (9) 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 (10) 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 to $T^{-1}\sum_{t=1}^{T}\theta_{1,t}$ in the computation of t_1.

For $r = 0$ and $r = 0.5$, 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 (i). This remains true even for instabilities that are large enough to be detected 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 magnitude of 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 out of 76 U.S. postwar macroeconomic time series on the 10% level using Andrews' (1993) QLR statistic. But based on Stock and Watson's (1998) method of obtaining median unbiased estimates for the magnitude h of a random walk instability by inverting the QLR 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- and three-parameter versions of a conditional consumption-based CAPM for 12 industry and 10 size sorted portfolios, using 7 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 macroeconomic application.

Also, as implied by Theorem 1 (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, in contrast to the modified statistic M_2. Comparing the power of M_1 and N_1, we find a moderate loss in power of the modified statistic.

For $r = 0.95$ these results change dramatically. While technically a stable VAR, the large autoregressive root leads to strong persistence in $(X_t, Z_t)'$. For such series, linearity of $T^{-1}\sum_{t=1}^{[\lambda T]} R_t R_t'$ in λ, i.e., Condition 1 (vii), is a bad approximation. Indeed, if one embeds $r = 0.95$ in a local-to-unity asymptotic framework (Chan and Wei (1987), Phillips (1987)) with local-to-unity parameter -5, one obtains an accurate description of the small sample behavior of the test statistics. But Condition 1 (vii) fails with $G_t(\theta_0)$ a function of local-to-unity processes. The results for $r = 0.95$ thus underline the crucial importance of Condition 1 (vii) for the conclusions of this paper. In empirical applications of Theorem 1, it is important to ensure that $T^{-1}\sum_{t=1}^{[\lambda T]} R_t R_t'$ is reasonably well approximated by a linear function.

By the linear algebra result discussed in Section 3, it follows that the results for t_2, $\mathrm{N}_{\mathrm{all}}$, N_1 and M_2 in Table 1 are independent of the correlation between the regressors ρ_{XZ}. Table 2 contains the results for t_1, N_2 and M_1 of the same Monte Carlo experiment for $\rho_{XZ} = 0$

Table 2: Small Sample Rejection Probabilities of 5% Nominal Tests in Percent

	break						random walk					
	$\rho_{XZ}=0$			$\rho_{XZ}=0.9$			$\rho_{XZ}=0$			$\rho_{XZ}=0.9$		
h	t_1	N_2	M_1	t_1	N_2	M_1	t_1	N_2	M_1	t_1	N_2	M_1
						$r=0$						
sm	5.9	53.8	54.5	6.2	53.8	14.0	5.8	44.1	44.6	6.3	44.1	15.4
lg	5.1	97.2	97.6	6.1	97.2	37.1	4.9	68.7	69.2	6.1	68.7	32.7
						$r=0.5$						
sm	6.7	50.4	49.7	6.9	50.4	13.0	7.0	42.5	42.5	6.9	42.5	14.8
lg	7.3	96.0	96.2	7.8	96.0	34.9	7.6	67.2	67.8	8.0	67.2	31.6
						$r=0.95$						
sm	13.3	20.6	15.9	12.6	20.6	6.0	15.4	23.1	19.9	14.5	23.1	9.2
lg	26.8	60.5	50.5	24.9	60.5	17.1	26.8	47.9	43.1	25.9	47.9	22.4

and $\rho_{XZ} = 0.9$. We find that the results for t_1 are not sensitive to the correlation between the regressors, the effect of instabilities in θ_1 on the usual Nyblom test N_2 for potential instabilities in θ_2 increases as ρ_{XZ} increases and, comparing the power of M_1 from Table 2 with the power of N_1 from Table 1, we find that the modified tests is substantially less powerful for strongly correlated regressors.

The second experiment studies the empirical relevance of the asymptotic results of Section 3 in a more applied context. Specifically, we consider the problem of conducting inference in a stylized model of monetary economics, whose baseline parameters are calibrated by estimates from real data. The two equation model consists of (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 (2005) for the theoretical derivation of this specification). Let π_t and s_t denote the inflation rate and unemployment gap at date t, respectively. The macroeconomic model underlying our simulation is given by the system

$$\Delta\pi_t = \phi E_t \Delta\pi_{t+1} + \kappa s_t + \varepsilon_t \tag{11}$$

$$s_t = \rho_{1,t} s_{t-1} + \rho_{2,t} s_{t-2} + \xi_t \tag{12}$$

where $\Delta\pi_t = \pi_t - \pi_{t-1}$, E_t is the conditional expectation at date t, and the disturbance terms ε_t and ξ_t are i.i.d. mean zero and multivariate normal. The NKPC (11) is expressed in first-differences rather than level to circumvent econometric problems generated by autoregressive roots close to unity (as illustrated in our OLS example above). The process of the driving variable s_t is specified as a simple AR(2) process. This is mainly for tractability since it allows us to derive a closed-form solution of the model that can be used to simulate data.

In addition to its tractability, the simple two equation system of (11) and (12) is an attractive example of our results, as 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 (11), 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.

We will focus on the forward solution of the two-equation system. Following Blanchard and Kahn (1980), the condition $\phi < 1$ guarantees a unique forward solution to (11). Under this condition and using the autoregressive process in (12), the unique reduced form of the two-equation system is

$$\begin{aligned} \Delta\pi_t &= \alpha_1 s_{t-1} + \alpha_2 s_{t-2} + (\varepsilon_t + \gamma\xi_t) \\ s_t &= \rho_1 s_{t-1} + \rho_2 s_{t-2} + \xi_t \end{aligned} \tag{13}$$

where

$$\alpha_1 = \frac{\kappa(\rho_1 + \phi\rho_2)}{1 - \phi\rho_1 - \phi^2\rho_2}, \ \alpha_2 = \frac{\kappa\rho_2}{1 - \phi\rho_1 - \phi^2\rho_2}, \text{ and } \gamma = \frac{\kappa}{1 - \phi\rho_1 - \phi^2\rho_2}. \tag{14}$$

We adopt the conventional 'anticipated utility' assumption in 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 (11) and (12) lead to time varying parameters of the reduced form parameters in (14), with the current values of α_1, α_2 and γ determined by the current values of ϕ, κ, 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 (11) is assumed stable throughout.

Leading the first equation in (13) one period and taking expectations conditional on information available at date $t-1$, the forecasting equation for $\Delta\pi_{t+1}$ becomes $\Delta\pi_{t+1} =$

$(\alpha_1\rho_1+\alpha_2)s_{t-1}+(\alpha_2\rho_2)s_{t-2}+(\varepsilon_{t+1}+\gamma\xi_{t+1}+\alpha\xi_t)$. 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 (11). The two equation system (11) and (12) 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 $g_t(\theta) = ((\Delta\pi_t - \phi\Delta\pi_{t+1} - \kappa s_t)s_{t-1}, (\Delta\pi_t - \phi\Delta\pi_{t+1} - \kappa s_t)s_{t-2}, (s_t - \rho_1 s_{t-1} - \rho_2 s_{t-2})s_{t-1}, (s_t - \rho_1 s_{t-1} - \rho_2 s_{t-2})s_{t-2})'$.

Since ε_t and ξ_t are i.i.d. multivariate Gaussian, it is straightforward to see that the stable reduced form model (13) satisfies Condition 2 (iii)-(v), so that Lemma 1 and standard arguments concerning stable GMM models yield Condition 1 (iv)–(vii) in an unstable model with parameter instabilities as specified in Condition 2 (i). Furthermore, since under the unstable model, $g_t(\theta_t) = (s_{t-1}\varepsilon_t, s_{t-2}\varepsilon_t, s_{t-1}\xi_t, s_{t-2}\xi_t)'$ is a martingale difference sequence, Lemma 2 and its discussion also 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.[2] For the NKPC (11), we use the full-sample estimates: $\phi = 0.73$ and $\kappa = -0.35$. For the AR(2) process of the unemployment gap (12), 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): changes in ρ_1 and ρ_2 are 0.48 and -0.18, respectively. The starting values of the AR(2) coefficients are $\rho_1 = 0.93$ and $\rho_2 = -0.43$, which are the estimates from the first sub-sample. Regarding the second moments of the disturbances, we obtain $E[\varepsilon_t^2] = 0.32$, $E[\xi_t^2] = 2.12$, and $E[\xi_t\varepsilon_t] = 0.05$ from full-sample estimation. We set the sample size in our experiment to $T = 160$.

We consider two forms of time varying paths for ρ_1 and ρ_2: a 'break' in the middle of the sample (as in our real-data estimation) $\rho_{1,t} = 0.93 + 6.1T^{-1/2}\mathbf{1}(t > T/2)$ and $\rho_{2,t} = -0.43 - 2.3T^{-1/2}\mathbf{1}(t > T/2)$; and a 'linear trend', representing a more gradual change in the

[2] $\Delta\pi_t$ 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 $s_t = u_t - \overline{u}_t$ where u_t is the unemployment rate and $\overline{u}_t$ is the natural rate of unemployment (NAIRU). The series u_t is obtained by converting a monthly series of unemployment for all workers (LHUR in DRI dataset) to the quarterly basis. The NAIRU series is constructed as a cubic spline in time, following Staiger, Stock and Watson (1997a, 1997b).

Table 3: Small Sample Rejection Probabilities of 5% Nominal Tests in Percent

t_ϕ	t_κ	$t_{\rho 1}$	$t_{\rho 2}$	N_{all}	N_ϕ	N_κ	$\mathrm{N}_{\phi\kappa}$	N_ρ	M_ϕ	M_κ	$\mathrm{M}_{\phi\kappa}$	M_ρ
all stable												
4.8	4.5	6.3	5.3	5.5	5.0	5.1	4.4	4.7	4.9	4.8	4.3	5.2
break												
4.3	4.0	8.0	7.9	53.7	15.9	22.4	21.9	95.5	4.3	5.3	5.3	90.6
linear trend												
3.9	6.0	7.6	8.1	46.6	13.4	18.8	16.2	86.8	4.2	5.4	5.5	72.9

coefficients $\rho_{1,t} = 0.93 + 6.1T^{-1/2}t/T$ and $\rho_{2,t} = -0.43 - 2.3T^{-1/2}t/T$. With this choice of instabilities, the AR(2) process is stationary throughout (the modulus of the largest root at the beginning of the sample is 0.66, and 0.78 at the end of the sample). A benchmark stable model sets (ρ_1, ρ_2) equal to the estimates of the first sub-sample. The initial values s_0 and $\Delta\pi_0$ are set to zero.

Table 3 reports empirical rejection probabilities of standard heteroskedasticity robust 5% level two-sided t-test of the four parameters ϕ, κ, ρ_1 and ρ_2 under the null hypothesis, standard 5% level Nyblom statistics N and modified Nyblom statistics M, defined in analogy to (9) and (10), that test the stability of all four parameters (N_{all}, which is equivalent to M_{all}), of ϕ, κ (N_ϕ, M_ϕ and N_κ, M_κ) and of (ϕ, κ) and (ρ_1, ρ_2) ($\mathrm{N}_{\phi\kappa}$, $\mathrm{M}_{\phi\kappa}$ and N_ρ, M_ρ). When ρ_i is time varying, the 'true' value of ρ_i is set to $T^{-1}\sum_1^T \rho_{i,t}$ in computing $t_{\rho i}$, $i = 1, 2$. All empirical rejection probabilities are based on asymptotic critical values, using 10,000 repetitions. From Table 3, it is clear that the specified magnitudes of the instabilities are not negligible in the sense of remaining undetected with high probability, even in an unspecific stability test of the four parameters. At the same time, the modified stability tests have close to nominal rejection probability when the subset of parameters under consideration is stable, as predicted by Theorem 1 (iii). The empirical rejection probabilities of the t-tests on the stable structural parameters ϕ and κ do not differ much from the nominal size, irrespective of the form of the instability, as predicted by Theorem 1 (i). On the other hand, the t-tests on ρ_i around the pseudo true value $T^{-1}\sum_1^T \rho_{i,t}$ for $i = 1, 2$ slightly overreject, although in this application, one would not necessarily be interested in conducting inference on this average.

Summarizing, the Monte Carlo experiment demonstrates that the asymptotic results of Theorem 1 approximate quite well the small-sample distributions of estimators and test statistics in the context of empirically relevant data generating processes.

5 Conclusion

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 quite general conditions, conventional GMM inference on parameters that ignores the instability remains asymptotically valid, as long as the instability is of moderate magnitude in the sense of not being detectable with probability one. 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 are a useful tool 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 one, even in the limit. In some instances, economic theory might be useful in making this choice, as in the Euler equation 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.

6 Appendix

The proofs of Lemmas 1—3, as well as Theorem 4, are based on the following Lemma.

Lemma 4 *If (i) $\psi : [0,1] \mapsto \mathbb{R}^r$ is a nonstochastic, bounded and piece-wise continuous function with at most a finite number of discontinuities; (ii) $T^{-1}\sum_{t=1}^{[\lambda T]} w_{T,t} \xrightarrow{p} \int_0^\lambda \vartheta(l)dl$ for all $0 \leq \lambda \leq 1$ and some nonstochastic Riemann-integrable function $\vartheta : [0,1] \mapsto \mathbb{R}^{r\times r}$ satisfying $\sup_{0\leq\lambda\leq 1} ||\vartheta(\lambda)|| < \infty$; (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) $T^{-1}\sum_{t=1}^{T} ||\tilde{w}_{T,t} - w_{T,t}|| \xrightarrow{p} 0$, 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 $T > T^*$, $P(||T^{-1}\sum_{t=1}^{[sT]} \tilde{w}_{T,t}\psi(t/T) - \int_0^s \vartheta(l)\psi(l)dl|| > \eta_1) < \eta_2$. 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 $P(\delta T^{-1}\sum_{t=1}^{T} ||w_{T,t}|| > \eta_1/4) < \eta_2/4$ for all $T > T_1^*$. Since ψ is continuous except at a finite number of points, it can be uniformly approximated by a sequence of step functions. There hence exists mutually disjoint intervals $\mathcal{I}_1, \cdots, \mathcal{I}_N$, $N < \infty$, satisfying $\bigcup_i \mathcal{I}_i = [0,1]$ and bounded vectors $c_1, \cdots, c_N$ 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

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

But

$$\begin{aligned} ||\int_0^s \vartheta(l)\varphi(l)dl - \int_0^s \vartheta(l)\psi(l)dl|| &\leq \delta\int_0^1 ||\vartheta(l)||dl \leq \eta_1/4 \\ ||T^{-1}\sum_{t=1}^{[sT]} w_{T,t}(\psi(t/T) - \varphi(t/T))|| &< \delta T^{-1}\sum_{t=1}^{T} ||w_{T,t}|| \\ ||T^{-1}\sum_{t=1}^{[sT]}(\tilde{w}_{T,t} - w_{T,t})\psi(t/T)|| &\leq \sup_{0\leq\lambda\leq 1} ||\psi(\lambda)|| \cdot T^{-1}\sum_{t=1}^{T} ||\tilde{w}_{T,t} - w_{T,t}|| \xrightarrow{p} 0 \end{aligned}$$

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

Now

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

and

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

If the ith 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

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

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

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

and $||T^{-1}\sum_{t/T\in\mathcal{I}_i} w_{T,t}-\int_{\mathcal{I}_i}\vartheta(l)dl|| \xrightarrow{p} 0$ 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

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

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

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

■

Proof of Lemma 1:

All following computations are under the stable model with density $\prod_{t=1}^{T} f_{T,t}(y_{T,t}, y_{T,t-1}, \cdots, y_{T,1}; \beta_0)$. The likelihood ratio statistic between the unstable model and the stable model 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

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

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

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

Therefore

$$\begin{aligned}\sum_{t=1}^{T}(\beta_t - \beta_0)'h_t(\tilde{\beta}_t)(\beta_t - \beta_0) &= T^{-1}\,\mathrm{tr}\sum_{t=1}^{T} h_t(\tilde{\beta}_t)B(t/T)B(t/T)' \\ \xrightarrow{p} -\mathrm{tr}\textstyle\int \Upsilon(l)B(l)B(l)'dl &= -\textstyle\int B(l)'\Upsilon(l)B(l)dl\end{aligned}$$

from a columnwise 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

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

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

$$\begin{aligned}T^{-1}\sum_{t=1}^{T} E[q_t^2|\mathfrak{F}_{t-1}] &= T^{-1}\sum_{t=1}^{T} B(t/T)'E[s_t(\beta_0)s_t(\beta_0)'|\mathfrak{F}_{t-1}]B(t/T) \\ &= \mathrm{tr}\,T^{-1}\sum_{t=1}^{T} E[s_t(\beta_0)s_t(\beta_0)'|\mathfrak{F}_{t-1}]B(t/T)B(t/T)' \\ \xrightarrow{p} \mathrm{tr}\textstyle\int \Upsilon(l)B(l)B(l)'dl &= \textstyle\int B(l)'\Upsilon(l)B(l)dl\end{aligned}$$

where the convergence in probability stems from a columnwise application of Lemma 4. By Corollary 3.1 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 Continuous Mapping Theorem (CMT), we conclude

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

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

Contiguity for Stochastic Parameter Paths:

Let B be random but independent of the data $\{y_{T,t}\}_{t=1}^T$ from the stable model for all T, and let B almost surely satisfy Condition 2 (i). Define $f_T(\{\beta_{T,t}\}_{t=1}^T) = \prod_{t=1}^T f_{T,t}(y_{T,t}, y_{T,t-1}, \cdots, y_{T,1}; \beta_{T,t})$, 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 measure of B and let A_T be the indicator function of a sequence of events with zero asymptotic probability in the stable model, i.e. $\int A_T f_T(\{\beta_0\}_{t=1}^T)d\mu_T \to 0$. By (one equivalent) definition of contiguity (see van der Vaart (1998), p. 87), we need to show that A_T has asymptotic probability zero also in the model with random parameter path $\{\beta_0 + T^{-1/2}B(t/T)\}_{t=1}^T$, i.e. $\int A_T E_B f_T(\{\beta_0 + T^{-1/2}B(t/T)\}_{t=1}^T)d\mu_T \to 0$. 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, $0 \leq \int A_T f_T(\{\beta_0 + T^{-1/2}b(t/T)\}_{t=1}^T)d\mu_T \leq 1$.

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 $T^{-1}\sum_{t=1}^T E[||g_t(\theta_t)||^{2+\epsilon}|\mathfrak{G}_{t-1}] = O_p(1)$ in the unstable model implies $T^{-1}\sum_{t=1}^T E[||g_t(\theta_t)||^2 \mathbf{1}[||g_t(\theta_t)|| > T^{1/2}a]|\mathfrak{G}_{t-1}] \xrightarrow{p} 0$ for all $0 < a < \infty$ in the unstable model. To invoke Corollary 2.7 of McLeish (1974) it thus remains to show that $T^{-1/2}\sup_{t \leq T}||g_t(\theta_t)|| \xrightarrow{p} 0$ 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 jth row of $\bar{G}_t$ is the jth row of $G_t(\cdot)$ evaluated at some θ on the line segment between θ_0 and θ_t.

We compute

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

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

$$\begin{aligned} T^{-1}\sup_{t\leq T}||\bar{G}_t|| &\leq pT^{-1}\sup_{t\leq T}\sup_{\theta\in\Theta_T}||G_t(\theta) - G_t(\theta_0) + G_t(\theta_0)|| \\ &\leq pT^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta_T}||G_t(\theta) - G_t(\theta_0)|| + pT^{-1}\sup_{t\leq T}||G_t(\theta_0)||. \end{aligned}$$

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

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

where

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}||\bar{G}_t(\theta_t - \theta_0)g_t(\theta_t)'|| &\leq (\sup_{0\leq\lambda\leq 1}||f(\lambda)||)T^{-1}\sum_{t=1}^{T}||\bar{G}_t||\cdot||T^{-1/2}g_t(\theta_t)|| \\ &\leq (\sup_{0\leq\lambda\leq 1}||f(\lambda)||)(T^{-1/2}\sup_{t\leq T}||g_t(\theta_t)||)T^{-1}\sum_{t=1}^{T}||\bar{G}_t|| \xrightarrow{p} 0 \end{aligned}$$

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

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}||\bar{G}_t|| &\leq pT^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta_T}||G_t(\theta) - G_t(\theta_0) + G_t(\theta_0)|| \\ &\leq pT^{-1}\sum_{t=1}^{T}\sup_{\theta\in\Theta_T}||G_t(\theta) - G_t(\theta_0)|| + pT^{-1}\sum_{t=1}^{T}||G_t(\theta_0)|| \end{aligned}$$

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

$$\begin{aligned} T^{-1}\sum_{t=1}^{T}||\bar{G}_t(\theta_t - \theta_0)(\theta_t - \theta_0)'\bar{G}_t'|| &\leq (\sup_{0\leq\lambda\leq 1}||f(\lambda)||)^2 T^{-2}\sum_{t=1}^{T}||\bar{G}_t||^2 \\ &\leq (\sup_{0\leq\lambda\leq 1}||f(\lambda)||)^2(T^{-1}\sup_{t\leq T}||\bar{G}_t||)T^{-1}\sum_{t=1}^{T}||\bar{G}_t|| \xrightarrow{p} 0. \end{aligned}$$

For the second claim of the Lemma, note that by contiguity, we have $||T^{-1}\sum_{t=1}^{[\lambda T]} g_t(\theta_t)g_t(\theta_t)' - \lambda V|| \xrightarrow{p} 0$ for each $0 \leq \lambda \leq 1$ in the unstable model, so that the result 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 jth row of $\tilde{h}_t$ is equal to the jth 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 Lemma 1 and 2 above,

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

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

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

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

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

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

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

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

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

and the result follows.

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 one, the first order condition of (2)

$$\left(T^{-1}\sum_{t=1}^{T} G_t(\hat{\theta})\right)' Q_T T^{-1/2}\sum_{t=1}^{T} g_t(\hat{\theta}) = 0 = \hat{\Gamma}' Q_T T^{-1/2}\sum_{t=1}^{T} g_t(\hat{\theta}) \tag{15}$$

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 one, 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 $t = 1, \cdots, [\lambda T]$ for $0 \leq \lambda \leq 1$

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

where the jth row of $\tilde{G}_t$ is the jth 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 a decreasing neighborhood $\mathcal{T}_T$ of θ_0 such that $P(\hat{\theta} \in \mathcal{T}_T) \to 1$. For T large enough to ensure that $\mathcal{T}_T \subset \Theta_0$

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

by Condition 1 (vi), so that by Condition 1 (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 of GMM (15), $||\hat{\Gamma} - \Gamma|| \xrightarrow{p} 0$ and $||Q_T - Q_0|| \xrightarrow{p} 0$ we find with these results that

$$T^{1/2}(\hat{\theta} - \theta_0) = \int f(l)dl - (\Gamma' Q_0 \Gamma)^{-1}\Gamma' Q_0 T^{-1/2}\sum_{t=1}^{T} g_t(\theta_t) + o_p(1). \tag{16}$$

The first result now follows from Condition 1 (iii) and the CMT. Since (16) implies $||\hat{\theta} - \theta_0|| = O_p(T^{-1/2})$, we have for all $0 \leq \lambda \leq 1$

$$T^{-1/2}\sum_{t=1}^{[\lambda T]} g_t(\hat{\theta}) = T^{-1/2}\sum_{t=1}^{[\lambda T]} g_t(\theta_t) + T^{1/2}\lambda\Gamma(\hat{\theta} - \theta_0) - \Gamma\int_0^\lambda f(l)dl + o_p(1). \tag{17}$$

Substituting (16) in (17) and rearranging yields

$$\begin{aligned} T^{-1/2}\sum_{t=1}^{[\lambda T]} g_t(\hat{\theta}) &= T^{-1/2}\sum_{t=1}^{[\lambda T]} g_t(\theta_t) - \lambda\Gamma(\Gamma' Q_0 \Gamma)^{-1}\Gamma' Q_0 T^{-1/2}\sum_{t=1}^{T} g_t(\theta_t) \\ &\quad -\Gamma\left(\int_0^\lambda f(l)dl - \lambda\int_0^1 f(l)dl\right) + R_T^*(\lambda) \end{aligned}$$

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 (vii) to uniform convergence over λ, $\sup_{\lambda\in[0,1]} ||T^{-1}\sum_{t=1}^{[\lambda T]}\tilde{G}_t - \lambda\Gamma|| \xrightarrow{p} 0$ and $\sup_{\lambda\in[0,1]} ||T^{-1}\sum_{t=1}^{[\lambda T]}\tilde{G}_t f(t/T) - \Gamma\int_0^\lambda f(l)dl|| \xrightarrow{p} 0$ 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.

References

Andrews, D. (1987): "Consistency in Nonlinear Econometric Models: A Generic Uniform Law of Large Numbers," *Econometrica*, 55, 1465–1471.

——— (1991): "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica*, 59, 817–858.

——— (1992): "Generic Uniform Convergence," *Econometric Theory*, 8, 241–257.

——— (1993): "Tests for Parameter Instability and Structural Change with Unknown Change Point," *Econometrica*, 61, 821–856.

——— (2003): "End-of-Sample Instability Tests," *Econometrica*, 71, 1661–1694.

Andrews, D., and W. Ploberger (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 P. Perron (1998): "Estimating and Testing Linear Models with Multiple Structural Changes," *Econometrica*, 66, 47–78.

Blanchard, O., and J. Gali (2005): "Real Wage Rigidities and the New Keynesian Model," *NBER working paper No. 1186.*

Blanchard, O., and C. Kahn (1980): "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48, 1305–1312.

Boivin, J. (1999): "Revisiting the Evidence on the Stability of Monetary VARs," *mimeo, Columbia University.*

Chan, N., and C. Wei (1987): "Asymptotic Inference for Nearly Nonstationary AR(1) Processes," *The Annals of Statistics*, 15, 1050–1063.

Cogley, T., and T. Sargent (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.* Oxford University Press, New York.

Elliott, G., and U. Müller (2005): "Efficient Tests for General Persistent Time Variation in Regression Coefficients," *mimeo, Princeton University. Forthcoming in Review of Economic Studies.*

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

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., A. Guay, and A. Hall (1997): "Predictive Tests for Structural Change with Unknown Breakpoint," *Journal of Econometrics*, 82, 209–233.

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

Hall, A., A. Inoue, and F. Peixe (2003): "Covariance Matrix Estimation and the Limiting Behavior of the Overidentifying Restrictions Test in the Presence of Neglected Structural Instability," *Econometric Theory*, 19, 962–983.

Hall, A., and A. Sen (1999): "Structural Stability Testing in Models Estimated by Generalized Method of Moments," *Journal of Business and Economic Statistics*, 17, 335–348.

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

Hansen, B. (1990): "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models," *working paper, Rochester University.*

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

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

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

Li, H. (2004): "The Econometrics of the Lucas Critique: Estimation and Testing of Euler Equation Models with Time-Varying Reduced-Form Coefficients," *mimeo, Princeton University.*

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

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

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

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

Newey, W., and K. West (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. (1987): “Towards a Unified Asymptotic Theory for Autoregression,” *Biometrika*, 74, 535–547.

PHILLIPS, P., AND W. PLOBERGER (1996): “An Asymptotic Theory of Bayesian Inference for Time Series,” *Econometrica*, 64, 381–412.

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

POLLARD, D. (2001): “Contiguity,” mimeo, Yale University.

PRIMICERI, G. (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., J. STOCK, AND M. WATSON (1997a): “How Precise are Estimates of the Natural Rate of Unemployment,” in *Reducing Inflation: Motivation and Strategy*, ed. by C. Romer, and D. Romer, pp. 195–242. University of Chicago Press.

——— (1997b): “The NAIRU, Unemployment, and Monetary Policy,” *Journal of Economic Perspectives*, 11, 33–51.

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

——— (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. (1998): *Asymptotic Statistics*. Cambridge University Press, Cambridge, UK.