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. 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.

JEL Classification: C32

Keywords: Structural Breaks, Parameter Stability Test, Contiguity, Euler Condition

*We thank two anonymous referees and the editor, Bernard Salanié, for very helpful comments and suggestions, as well as Graham Elliott and Mark Watson and 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 helpful discussions. Müller gratefully acknowledges financial support by the NSF via 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 (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 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.

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 macroeconomic 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 one in the limit—see Elliott and Müller (2007) for some quantitative results on break date confidence sets.
In this paper, we analyze models where only a subset of parameters are 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

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

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 \( \theta_2 \) 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 \( \tilde{Z}_t \), \( (1) \) becomes

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

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 \( \varepsilon_t \) dominates the randomness of the OLS estimator of \( \theta_2 \), and inference on \( \theta_2 \) remains largely unaffected by the instability of \( \theta_1 \). 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 \( \theta_2 \), 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 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. 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

\[ \Delta \pi_t = \phi E_t \Delta \pi_{t+1} + \kappa s_t + \varepsilon_t \]  

\[ s_t = \rho_{1,t}s_{t-1} + \rho_{2,t}s_{t-2} + \xi_t \]

where \( \pi_t \) and \( s_t \) are the inflation rate and unemployment gap, respectively, \( E_t \) is the conditional expectation at date \( t \), and the disturbance terms \( \varepsilon_t \) and \( \xi_t \) are i.i.d. mean zero. In this example, economic theory has direct implications for the stability of the various parameters: the coefficients \( \rho_1 \) and \( \rho_2 \) are functions of current monetary policy. With a time varying monetary policy, \( \rho_1 \) and \( \rho_2 \) 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 \( \phi \) and \( \kappa \) to be stable, even in the face of a time varying monetary policy. Our results imply that one might estimate the system (3) and (4) by standard GMM ignoring the time variability in \( \rho_1 \) and \( \rho_2 \), and one still obtains approximately valid inference about the structural parameters \( \phi \) and \( \kappa \), at least as long as the time variability in \( \rho_1 \) and \( \rho_2 \) is not too pronounced. We return to this example in Section 4 below 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 $T^{-1/2}$ neighborhood.

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,\footnote{As noted above, inference based on Bai and Perron’s (1998) results face the same problem in reverse— their approach yields poor approximations for local parameter instabilities. See Elliott and Müller (2007) for details.} where, for instance, $\theta_{1,t} - \theta_{1,0} = 1[t \geq T/2]$ rather than $\theta_{1,t} - \theta_{1,0} = T^{-1/2}1[t \geq T/2]$ in (1).\footnote{For such a fixed parameter instability $\theta_{1,t} - \theta_{1,0} = 1[t \geq T/2]$, 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 $\theta_2$ 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.}

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 $\epsilon > 0$, and to never reject otherwise. Since the p-value of a good parameter stability test converges to zero in probability for instabilities that are larger than the $T^{-1/2}$ neighborhood, 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 $\epsilon$ too small relative to the actual sample size might still yield poor approximations, so that a judgement must still be made. In Section 3 below, 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 macroeconomic 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 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 (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: Subsection 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 subsection 2.2, followed by a discussion of its implications for econometric practice. In subsection 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 an Appendix.

2 Asymptotic Results

2.1 Model and High-Level Condition

Consider a GMM model with the unknown \(m \times 1\) parameter vector \(\theta\), 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 \((\Omega, \mathcal{G}, P)\), 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 \(\theta\).

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 \(\theta_0\) 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

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

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 \(g(y_t, \theta)\). All limits are taken as \(T \to \infty\). We write ‘\(\mathcal{P}\)’ 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

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

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

(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, and left and right limits everywhere.
(ii) In some neighborhood $\Theta_0$ of $\theta_0$, $g_t(\theta)$ is differentiable in $\theta$ 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) $\hat{\theta} \overset{p}{\rightarrow} \theta_0$.

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

(vi) $T^{-1} \sum_{t=1}^{T} ||G_t(\theta_0)|| = O_p(1)$, $T^{-1} \sup_{t \leq T} ||G_t(\theta_0)|| \overset{p}{\rightarrow} 0$ and for any decreasing neighborhood $\Theta_T$ of $\theta_0$ contained in $\Theta_0$, i.e. $\Theta_T = \{ \theta : ||\theta - \theta_0|| < c_T \} \subset \Theta_0$ for some sequence of real numbers $c_T \rightarrow 0$, $T^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta_T} ||G_t(\theta) - G_t(\theta_0)|| \overset{p}{\rightarrow} 0$.

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

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 (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 $T^{-1} \sum_{t=1}^{T} G_t(\theta_0)$ in $\lambda$ as imposed in part (vii) (which, given the condition in part (vi), is equivalent to the approximate linearity of $T^{-1} \sum_{t=1}^{T} G_t(\theta_0)$). This assumption entails that averages of $G_t(\theta_0)$ are approximately equal to $\Gamma$ 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 $T^{-1} \sum_{t=1}^{T} g_t(\theta_0) = O_p(1)$.

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 $\{g_t(\theta_t)\}_{t=1}^{T}$ need to be checked in the unstable model, a complication to which we return in Section 2.4 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 $T^{-1} \sum_{t=1}^{T} g_t(\theta) = N(0, I_p)\Gamma$. Furthermore, if in addition, $\sup_{\lambda \in [0,1]} \|T^{-1} \sum_{t=1}^{T} G_t(\theta_0) - \lambda \Gamma\| \overset{p}{\to} 0$ and $T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t) \Rightarrow V^{1/2}W(\cdot)$ with $W$ a $p \times 1$ standard Wiener process, then

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

(ii) $T^{-1/2} \sum_{t=1}^{T} g_t(\hat{\theta}) = N(0, (I_p - \Gamma(\Gamma'Q_0\Gamma)^{-1}\Gamma'Q_0)\Gamma(\Gamma'Q_0\Gamma)^{-1}\Gamma'Q_0')V(I_p - \Gamma(\Gamma'Q_0\Gamma)^{-1}\Gamma'Q_0')\Gamma$.

where $\hat{\Sigma}_{\theta} = (\hat{\Gamma}'Q_T\hat{\Gamma})^{-1}\hat{\Gamma}'Q_T\hat{\Gamma}(\hat{\Gamma}'Q_T\hat{\Gamma})^{-1}$ and $\hat{\Gamma} = T^{-1} \sum_{t=1}^{T} G_t(\hat{\theta}) \overset{p}{\to} \Gamma$. Furthermore, $\sup_{\lambda \in [0,1]} \|T^{-1} \sum_{t=1}^{T} G_t(\theta_0) - \lambda \Gamma\| \overset{p}{\to} 0$ and $T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t) \Rightarrow V^{1/2}W(\cdot)$ with $W$ a $p \times 1$ standard Wiener process, then
Part (i) of Theorem 1 shows that standard asymptotically Gaussian inference based on \( \hat{\theta} \) and \( \Sigma_\theta \) remains valid for the stable subset of the parameters (where \( \theta_t \) is the same for all \( t \) and equal to \( \theta_0 \) 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, \( ||\hat{\theta} - T^{-1} \sum_{t=1}^{T} \theta_t|| \overset{p}{\to} 0 \). 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 (6)

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

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 \( \theta_t \) and \( \hat{\theta} \). Standard arguments imply that under Condition 1, \( T^{-1} \sum_{t=1}^{T} \tilde{G}_t \overset{p}{\to} \Gamma \). 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 \( \Gamma \) 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: \( T^{-1} \sum_{t=1}^{T} \tilde{G}_t f(t/T) \overset{p}{\to} \lim_{T \to \infty} \Gamma T^{-1} \sum_{t=1}^{T} f(t/T) \). 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 (7) and taking limits, revealing the relevance of this result for conducting asymptotically valid inference in partially stable models.

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.

2.3 Stability Tests in Partially Unstable Models

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 (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 \( T^{-1/2} \sum_{t=1}^{[T]} g_t(\hat{\theta}) \). 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 \( \theta \). 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 \( Q_T = \hat{\Gamma}_T^{-1} \).

One might invoke the analysis of Sowell (1996), who derives tests of

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

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

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

of \( \hat{\Gamma}_T^{-1/2}T^{-1/2} \sum_{t=1}^{[T]} g_t(\hat{\theta}) \Rightarrow J(\cdot) \) is asymptotically sufficient for the local alternative \( f \), where \( \Sigma_{\hat{\theta}} = (\Gamma'V^{-1}\Gamma)^{-1} \) 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 (8) that maximize power against alternatives where only the first \( m_0 \) elements of \( \theta \) are time
varying may be based on functionals of

\[(C^\prime \hat{\Sigma}_{\theta}^{-1}C)^{-1/2}C^\prime \hat{\Gamma}' \hat{V}_T^{-1}T^{-1/2} \sum_{t=1}^{[T]} g_t(\hat{\theta}) \Rightarrow W_{m_0}(\cdot) - \cdot W_{m_0}(1)\]

\[(C^\prime \Sigma_{\theta}^{-1}C)^{-1/2}C^\prime \Sigma_{\theta}^{-1/2}W_m.\]

In general, as long as \(\Sigma_{\theta}\) 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 \(\theta\) 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

\[H_0 : C^\prime \theta_t \text{ is constant in } t \quad \text{against} \quad H_1 : C^\prime \theta_t \text{ depends on } t\]

which allows for local instabilities of the last \(m - m_0\) parameters in \(\theta\) under the null hypothesis.

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

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

\[= \tilde{W}_{m_0}(\cdot) - \cdot \tilde{W}_{m_0}(1) + (C^\prime \Sigma_{\theta}C)^{-1/2}C^\prime \left(\int_0^1 f(l)dl - \cdot \int_0^1 f(l)dl\right),\]

where \(\tilde{W}_{m_0} = (C^\prime \Sigma_{\theta}C)^{-1/2}C^\prime \Sigma_{\theta}^{1/2}W_m.\) The asymptotic null distribution of stability tests based on (11) does not depend on the local instabilities in the last \(m - m_0\) elements of \(\theta\) because \(C^\prime f\) is equal to zero whenever the first \(m_0\) elements of \(f\) are. More formally, the right-hand side of (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 \(\theta\), so that efficient tests based on (11) yield best invariant tests of (8) based on \(J.\)

\[3^\text{The results of Müller (2007) provide a sense in which this procedure is asymptotically efficient for the original GMM problem, rather than only for the limiting process } J.\]
When one applies the same type of test statistic to (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 left-hand side of (9) and (11), one obtains the same asymptotic distribution when all parameters are stable, and thus the same critical value. But in contrast to (9), under the null hypothesis (10) of stability of the first $m_0$ elements of $\theta$, the second summand in (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 (11) will not be affected by instabilities in the non-tested parameters.\textsuperscript{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 (11) typically have lower power than tests based on (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 $\Sigma_\theta$ 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 (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 (11). For tests that seek to detect potential instabilities in all parameters, i.e. $C = I_m$, (11) reduces to (9).

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 (11), whose asymptotic rejection probabilities are a function of the stability of the parameters under consideration only.

\textsuperscript{4}One might worry about the local nature of the robustness against instabilities in the last $m - m_0$ elements of $\theta$ 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 $\theta$ is smaller than some threshold $\epsilon > 0$, similar to the discussion in the introduction.
2.4 Justifying Condition 1 via Contiguity

As noted in Section 2.1 above, 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 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.

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 \( f = 0 \)). The formal argument relies on the concept of contiguity.

**Definition 2** 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 \( X_T \overset{p}{\rightarrow} 0 \) under \( P_T^2 \) implies \( X_T \overset{p}{\rightarrow} 0 \) 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 model can be described by the evolution of the \( k \times 1 \) parameter \( \beta, k \geq p \), 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 \( \beta \) 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 \( \theta \). 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 \( \mathfrak{H}_{T,t} \) be the \( \sigma \)-field generated by \( \{y_{T,s}\}_{s=1}^t \) and \( \mathfrak{H}_{T,0} \) be the trivial \( \sigma \)-field. We again omit the dependence on \( T \) of \( \beta_{T,t}, s_{T,t}, h_{T,t} \) and \( \mathfrak{H}_{T,t} \) 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 \( \beta_t \) 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, and left and right limits everywhere.

(ii) In some neighborhood \( \mathcal{B}_0 \) of \( \beta_0 \), \( l_t(\beta) \) is twice differentiable a.s. with respect to \( \beta \) for \( t = 1, \cdots, T \).

Furthermore, in the stable model,

(iii) \( \{s_t(\beta_0), \mathfrak{H}_t\} \) is a square-integrable martingale difference array with \( T^{-1} \sum_{t=1}^{\lfloor \lambda T \rfloor} E[s_t(\beta_0)s_t(\beta_0)'/\mathfrak{H}_{t-1}] \overset{p}{\rightarrow} \int_0^1 \mathcal{Y}(l)dl \) for all \( 0 \leq \lambda \leq 1 \) and some non-stochastic bounded Riemann integrable matrix function \( \mathcal{Y} : [0,1] \mapsto \mathbb{R}^{k \times k} \),

\( T^{-1} \sup_{t \leq T} \|E[s_t(\beta_0)s_t(\beta_0)'|\mathfrak{H}_{t-1}]\| \overset{p}{\rightarrow} 0 \) and there exists \( \nu > 0 \) such that \( T^{-1} \sum_{t=1}^T E[\|s_t(\beta_0)\|^{2+\nu}|\mathfrak{H}_{t-1}] = O_p(1) \).

(iv) \( T^{-1} \sum_{t=1}^T \|h_t(\beta_0)\| = O_p(1), T^{-1} \sup_{t \leq T} \|h_t(\beta_0)\| \overset{p}{\rightarrow} 0 \) and for any decreasing neighborhood \( \mathcal{B}_T \) of \( \beta_0 \) contained in \( \mathcal{B}_0 \), \( T^{-1} \sum_{t=1}^T \sup_{\beta \in \mathcal{B}_T} \|h_t(\beta) - h_t(\beta_0)\| \overset{p}{\rightarrow} 0 \).
(v) For all \( 0 \leq \lambda \leq 1 \), \( T^{-1} \sum_{t=1}^{[NT]} h_t(\beta_0) \xrightarrow{p} -\int_0^\lambda \Upsilon(l) dl \).

Part (i) makes the same assumption on the form of the instability in \( \beta \) as Condition 1 (i) does on \( \theta \). 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, \( \{s_t(\beta_0), \mathfrak{g}_t\} \) and \( \{s_t(\beta_0)s_t(\beta_0)' + h_t(\beta_0), \mathfrak{g}_t\} \) are martingale difference arrays by construction—see Hall and Heyde (1980), Chapter 6.2. The matrix function \( \Upsilon \) represents the average rate of (conditional) information accrual on the time scale of the the sample fraction. For stationary stable models, \( \Upsilon \) is constant and equal to the probability limit of \( (-T^{-1} \sum_{t=1}^{T} h_t(\beta_0)) \) and \( T^{-1} \sum_{t=1}^{T} E[s_t(\beta_0)s_t(\beta_0)'|\mathfrak{g}_{t-1}] \). The point-wise convergences in \( \lambda \) 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.

Under and 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 (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 $\mathcal{G}_{T,t}$, where $g_{T,s}(\theta_{T,s})$ is measurable with respect to $\mathcal{G}_{T,t}$ 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, $\{g_t(\theta_t), \mathcal{G}_t\}_{t=1}^T$ is a martingale difference array and there exists $\zeta > 0$ such that $T^{-1} \sum_{t=1}^T E[||g_t(\theta_t)||^{2+\zeta}|\mathcal{G}_{t-1}] = O_p(1)$, and in the stable model, Condition 1 parts (i),(ii),(vi) and (vii) hold, $T^{-1/2} \sup_{t \leq T} ||g_t(\theta_0)|| \overset{p}{\to} 0$ and $T^{-1} \sum_{t=1}^T g_t(\theta_0)g_t(\theta_0)' \overset{p}{\to} V$ 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 $T^{-1} \sum_{t=1}^{[\lambda T]} g_t(\theta_0)g_t(\theta_0)' \overset{p}{\to} \lambda V$ for all $0 \leq \lambda \leq 1$ in the stable model, then $T^{-1/2} \sum_{t=1}^T g_t(\theta_t) \Rightarrow V^{1/2}W(\cdot)$ 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 $\{g_t(\theta_t), \mathcal{G}_t\}_{t=1}^T$ 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 $\mathcal{G}_t$ and $E[||\varepsilon_t||^{2+\zeta}|\mathcal{G}_{t-1}] \leq \tilde{M}_\varepsilon$ a.s under the unstable model. Then $T^{-1} \sum_{t=1}^T E[||g_t(\theta_t)||^{2+\zeta}|\mathcal{G}_{t-1}] \leq \tilde{M}_\varepsilon T^{-1} \sum_{t=1}^T ||x_{t-1}||^{2+\zeta}$ a.s. in the unstable model, and it suffices to show that $T^{-1} \sum_{t=1}^T ||x_{t-1}||^{2+\zeta} = O_p(1)$ 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 $||T^{-1/2} \sum_{t=1}^T g_t(\theta_0) - T^{-1/2} F' \sum_{t=1}^T s_t(\beta_0)|| \overset{p}{\to} 0$ 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 $V = F' \int \bar{Y}(s)ds F$.

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. 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 one 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 New Keynesian Phillips Curve example introduced in the introduction. We calibrate the data generating process to quarterly US 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

\[ Y_t = X_t \theta_{1,t} + Z_t \theta_2 + \theta_3 + \epsilon_t, \quad t = 1, \ldots, T \]  

(12)

where \( \epsilon_t \sim i.i.d. \mathcal{N}(0,1) \) and \((X_t, Z_t)'\) is a zero-mean stationary Gaussian VAR(1) independent of \(\{\epsilon_t\}\) 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 \(\hat{\epsilon}_t\) 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 \( \hat{V}_T = T^{-1} \sum_{t=1}^{T} R_t R_t' \hat{\varepsilon}_t^2 \). The parameter \( r \) governs the degree of autoregressive persistence in the nonconstant 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 \( \theta_2 \), \( C = (0, 1, 0)' \). With \( \hat{\Sigma}_\theta = (\hat{\Gamma}'\hat{V}_T^{-1}\hat{\Gamma})^{-1} \), the non-modified Nyblom statistic based on (9) is then given by

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

and the modified Nyblom statistic based on (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 (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 \( \theta_1 \): a ‘break’ in the middle of the sample, \( \theta_{1,t} = hT^{-1/2}[t > T/2] \); 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 table) 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 table) 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 \( \theta_1 \) and \( \theta_2 \) (\( t_1 \) and \( t_2 \)) under the null hypothesis, of the usual Nyblom statistics (13) for the constancy of all three coefficients (\( N_{all} \)) and of \( \theta_1 \) and \( \theta_2 \) (\( N_1 \) and \( N_2 \)), and of the modified Nyblom statistics (14) for the constancy of the coefficients \( \theta_1 \) and \( \theta_2 \) (\( M_1 \) 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 \( \theta_1 \) is set to \( T^{-1} \sum_{t=1}^{T} \theta_{1,t} \) in the computation of \( t_1 \). By inspection of (13) and (14), note that \( M_2 \) and \( N_1 \) are invariant to the transformation

---

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

<< Table 1 about here >>

The empirical rejection probability of the t-test on the stable coefficient $\theta_2$ is very little affected by the instability in $\theta_1$, as predicted by Theorem 1 (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 $N_{\text{all}}$ 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 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 $\theta_1$ 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 $\theta_2$, $N_2$, is strongly affected by the instability in $\theta_1$ 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 $\rho_{XZ} = 0.9$. 

The results 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 $\theta_2$ as a function of $r \in [0, 1)$. While technically a stationary VAR, for large autoregressive roots $r < 1$, linearity of $T^{-1}\sum_{t=1}^{[XT]} R_t R'_t$ in $\lambda$, i.e., Condition 1 (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 New Keynesian Phillips Curve Model

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

\begin{align}
\Delta \pi_t &= \alpha_1 s_{t-1} + \alpha_2 s_{t-2} + \nu_t \\
s_t &= \rho_1 s_{t-1} + \rho_2 s_{t-2} + \xi_t
\end{align}

where $\nu_t = \varepsilon_t + \gamma \xi_t$ and

\begin{align}
\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}, & \gamma &= \frac{\kappa}{1 - \phi \rho_1 - \phi^2 \rho_2}.
\end{align}

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 (16), with the current values of $\alpha_1$, $\alpha_2$ and $\gamma$ determined by the current values of $\rho_1$ and $\rho_2$. Note that due to the interaction via the expected future inflation term, instabilities in $\rho_1$ and $\rho_2$ lead to unstable reduced form parameters $\alpha_1$ and $\alpha_2$, 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 $\Delta \pi_{t+1} = (\alpha_1 \rho_1 + \alpha_2) s_{t-1} + \alpha_1 \rho_2 s_{t-2} + \nu_{t+1} + \alpha_1 \xi_t$. 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 (3). The two equation system (3) and (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 (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, under the unstable model, $g_t(\theta_t) = (s_{t-1} (\varepsilon_t - \phi \nu_{t+1}), s_{t-2} (\varepsilon_t - \phi \nu_{t+1}), s_{t-1} \xi_t, s_{t-2} \xi_t)'$, so that $g_t(\theta_t)$ is a vector moving average process of order one, and arguments similar to those applied in the proof of Lemma 2 yield $T^{-1/2} \sum_{t=1}^{[T]} g_t(\theta_t) \Rightarrow V^{1/2} W(\cdot)$ 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.\textsuperscript{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 $\rho_1$ and $\rho_2$ 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 $\rho_1$ and $\rho_2$ that undergo a discrete shift in the middle of the sample, whose magnitude relative to this estimate is described by $K$, i.e.

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

\textsuperscript{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 $s_t = u_t - \overline{u}_t$ 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).
Under (17), the AR(2) process is stationary throughout when \(0 \leq K \leq 2.5\). Regarding the second moments of the disturbances, we determine \(E[\nu_t^2] = 0.55\), \(E[\xi_t^2] = 2.11\), and \(E[\xi_t \nu_t] = -0.01\) 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 zero, and \(T = 160\).

The left panel of Figure 2 depicts the rejection probability of 5% nominal level t-tests \(t_\phi\) and \(t_\kappa\) on the stable structural parameters \(\phi\) and \(\kappa\) as a function of \(K\), based on the usual GMM 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 (13) and (14), that test the stability of \(\phi\) and \(\kappa\). 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 \(\rho_1\) and \(\rho_2\) 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 (i), the null rejection probabilities of \(t_\phi\) and \(t_\kappa\) 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 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 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.

5 Appendix

The proofs of Theorem 1 and Lemmas 1-3 are based on the following Lemma.

Lemma 4 If (i) \( \psi : [0, 1] \rightarrow \mathbb{R}^d \) is a nonstochastic, 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 \( T^{-1} \sum_{t=1}^{[sT]} w_{T,t} \xrightarrow{p} \int_0^s \vartheta(l)dl \) for all \( 0 \leq \lambda \leq 1 \) and some nonstochastic Riemann-integrable function \( \vartheta : [0, 1] \rightarrow \mathbb{R}^{d \times d} \) 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 \( T > T^* \),

\[
P(||T^{-1} \sum_{t=1}^{[sT]} \tilde{w}_{T,t} \psi(t/T) - \int_0^s \vartheta(l)\psi(l)dl|| > \eta_1) < \eta_2.
\]

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 continuity on a compact interval implies uniform continuity on that interval, \( \psi \) is continuous except at a finite number of points and \( \psi \) has left and right limits everywhere, the function \( \psi \) can be uniformly approximated by a sequence of step functions. There hence exists mutually disjoint intervals \( I_1, \ldots, I_N, N < \infty \), satisfying \( \bigcup_i I_i = [0, 1] \) and bounded vectors \( c_1, \ldots, c_N \) such that \( \varphi(\lambda) = \sum_{i=1}^{N} 1[\lambda \in I_i]c_i \) and \( \sup_{0 \leq \lambda \leq 1} ||\psi(\lambda) - \varphi(\lambda)|| < \delta \). We have

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

But

\[
||\int_0^{s} \vartheta(l) \varphi(l) dl - \int_0^{s} \vartheta(l) \psi(l) dl|| \leq \delta \int_0^{1} ||\vartheta(l)|| dl \leq \eta_1/4
\]

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

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

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

\[
T^{-1} \sum_{t=1}^{[sT]} w_{T,t} \varphi(t/T) = T^{-1} \sum_{t=1}^{[sT]} w_{T,t} \sum_{i=1}^{N} 1[t/T \in I_i]c_i
\]

\[
= \sum_{i=1}^{N} T^{-1} (\sum_{t \leq [sT], t/T \in I_i} w_{T,t}) c_i
\]

and

\[
||\sum_{i=1}^{N} (T^{-1} \sum_{t \leq [sT], t/T \in I_i} w_{T,t}) c_i - \sum_{i=1}^{N} (\int_{I_i} 1[l \leq s] \vartheta(l) dl) c_i||
\]

\[
\leq \sup_{i \leq N} ||c_i|| \cdot \sum_{i=1}^{N} \sum_{t \leq [sT], t/T \in I_i} ||w_{T,t} - \int_{I_i} 1[l \leq s] \vartheta(l) dl||.
\]
If the \( i \)th interval is of the form \( \mathcal{I}_i = (a_i, b_i) \) then \( T^{-1} \sum_{t/T \in \mathcal{I}_i} w_{T,t} = T^{-1} \sum_{t=[a_i T]+1}^{[b_i T]} w_{T,t} \) and hence

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

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

\[
||T^{-1} \sum_{t/T \in \mathcal{I}_i} w_{T,t}|| \leq ||T^{-1} \sum_{t=[a_i T]+1}^{[b_i T]} w_{T,t}|| + T^{-1}||w_{T,[a_i T]}|| + T^{-1}||w_{T,[b_i T]}||
\]

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

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

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

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

\[\blacksquare\]

**Proof of Theorem 1:**

Since \( g \) is differentiable on \( \Theta_0 \), and \( \hat{\theta} \overset{p}{\rightarrow} \theta_0 \), for large enough \( T \) and with probability converging to one, the first order condition of (6)

\[
\left(T^{-1} \sum_{t=1}^{T} G_t(\hat{\theta})\right)' Q_T T^{-1/2} \sum_{t=1}^{T} g_t(\hat{\theta}) = 0 = \hat{\varphi}' Q_T T^{-1/2} \sum_{t=1}^{T} g_t(\hat{\theta})
\]

is satisfied. Also, since \( \hat{\theta} \overset{p}{\rightarrow} \theta_0 \) and \( ||\theta_t - \theta_0|| \rightarrow 0 \), for large enough \( T \) and with probability converging to one, all line segments between \( \hat{\theta} \) and \( \theta_t \) are subsets of \( \Theta_0 \). 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 \( 0 \leq \lambda \leq 1 \)

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

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

(18)
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 \( \theta_t \) and \( \hat{\theta} \).

Since \( \hat{\theta} \xrightarrow{p} \theta_0 \), there exists a decreasing neighborhood \( T_T \) of \( \theta_0 \) such that \( P(\hat{\theta} \in T_T) \to 1 \). For \( T \) large enough to ensure that \( T_T \subset \Theta_0 \), with probability converging to one,

\[
T^{-1} \sum_{t=1}^{[NT]} \| \tilde{G}_t - G_t(\theta_0) \| \leq pT^{-1} \sum_{t=1}^{T} \sup_{\theta \in T_T} \| G_t(\theta) - G_t(\theta_0) \| \xrightarrow{p} 0
\]

by Condition 1 (vi), so that by Condition 1 (vii), \( T^{-1} \sum_{t=1}^{[NT]} \tilde{G}_t \xrightarrow{p} \lambda \Gamma \) for all \( 0 \leq \lambda \leq 1 \). Also, we can apply Lemma 4 to \( T^{-1} \sum_{t=1}^{[NT]} \tilde{G}_tf(t/T) \) and find \( T^{-1} \sum_{t=1}^{[NT]} \tilde{G}_tf(t/T) \xrightarrow{p} \Gamma \int_0^\lambda f(l)dl \) for all \( 0 \leq \lambda \leq 1 \). From the first order condition of GMM (18), \( \| \Gamma - \Gamma' \| \xrightarrow{p} 0 \) and \( \| Q_T - Q_0 \| \xrightarrow{p} 0 \) we find with these results that

\[
T^{1/2}(\hat{\theta} - \theta_0) = \int f(l)dl - (\Gamma'Q_0\Gamma)^{-1}\Gamma'Q_0 T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t) + o_p(1). \tag{19}
\]

The first result now follows from Condition 1 (iii) and the CMT. Since (19) implies \( \| \hat{\theta} - \theta_0 \| = O_p(T^{-1/2}) \), we have for all \( 0 \leq \lambda \leq 1 \)

\[
T^{-1/2} \sum_{t=1}^{[NT]} g_t(\hat{\theta}) = T^{-1/2} \sum_{t=1}^{[NT]} g_t(\theta_t) + T^{1/2}\lambda \Gamma(\hat{\theta} - \theta_0) - \Gamma \int_0^\lambda f(l)dl + o_p(1). \tag{20}
\]

Substituting (19) in (20) and rearranging yields

\[
T^{-1/2} \sum_{t=1}^{[NT]} g_t(\hat{\theta}) = T^{-1/2} \sum_{t=1}^{[NT]} g_t(\theta_t) - \lambda \Gamma(\Gamma'Q_0\Gamma)^{-1}\Gamma'Q_0 T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t) - \Gamma \left( \int_0^\lambda f(l)dl - \lambda \int_0^1 f(l)dl \right) + R^*_T(\lambda)
\]

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 (vii) to uniform convergence over \( \lambda \), \( \sup_{\lambda \in [0,1]} \| T^{-1} \sum_{t=1}^{[NT]} \tilde{G}_t - \lambda \Gamma \| \xrightarrow{p} 0 \) and \( \sup_{\lambda \in [0,1]} \| T^{-1} \sum_{t=1}^{[NT]} \tilde{G}_tf(t/T) - \Gamma \int_0^\lambda f(l)dl \| \xrightarrow{p} 0 \) 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 \( T^{-1/2} \sum_{t=1}^{T} g_t(\theta_t) \Rightarrow V^{1/2}W(\cdot) \) and the CMT.

**Proof of Lemma 1:**

All of 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 model
and the stable model is \( LR_T = \exp \left[ \sum_{t=1}^{T} (l_t(\beta_t) - l_t(\beta_0)) \right] \). Let \( B_T = \{ \beta : ||\beta - \beta_0|| \leq T^{-1/2} \sup_{0 \leq \lambda \leq 1} ||B(\lambda)|| \} \). For \( T \) large enough to ensure that \( B_T \subset B_0 \), 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 \( \beta_0 \) and \( \beta_t \). From Condition 2 (iv),

\[
T^{-1} \sum_{t=1}^{T} ||h_t(\tilde{\beta}_t) - h_t(\beta_0)|| \leq T^{-1} \sum_{t=1}^{T} \sup_{\beta \in B_T} ||h_t(\beta) - h_t(\beta_0)|| \xrightarrow{p} 0.
\]

Therefore

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

\[
\xrightarrow{p} - \text{tr} \int \Upsilon(l)B(l)B(l)dl = -\int B(l)\Upsilon(l)B(l)dl
\]

from a columnwise application of Lemma 4.

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

\[
T^{-1} \sum_{t=1}^{T} E[|q_t|^{2+\nu}|\tilde{\beta}_t-1] \leq \sup_{0 \leq \lambda \leq 1} ||B(\lambda)||^{2+\nu} T^{-1} \sum_{t=1}^{T} E[||s_t(\beta_0)||^{2+\nu}|\tilde{\beta}_t-1]
\]

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

\[
T^{-1} \sum_{t=1}^{T} E[q_t^2|\tilde{\beta}_t-1] = T^{-1} \sum_{t=1}^{T} B(t/T)'E[s_t(\beta_0)s_t(\beta_0)'|\tilde{\beta}_t-1]B(t/T)
\]

\[
= \text{tr} T^{-1} \sum_{t=1}^{T} E[s_t(\beta_0)s_t(\beta_0)'|\tilde{\beta}_t-1]B(t/T)B(t/T)'
\]

\[
\xrightarrow{p} \text{tr} \int \Upsilon(l)B(l)B(l)dl = \int B(l)\Upsilon(l)B(l)dl
\]

where the convergence in probability stems from a columnwise 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 Continuous Mapping Theorem (CMT), we conclude

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

**Contiguity for Stochastic Parameter Paths:**

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 $\sigma$-finite measure $\mu_T$, 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. $\int A_T E_B f_T(\{\beta_0\}_{t=1}^T) d\mu_T \to 0$. 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 $\int A_T E_B f_T(\{\beta_0 + T^{-1/2}B(t/T)\}_{t=1}^T) d\mu_T \to 0$, which follows from $\int A_T E_B f_T(\{\beta_0 + T^{-1/2}b(t/T)\}_{t=1}^T) d\mu_T \to 0$ for almost all realizations $B = b$ by Lemma 1 and the dominated convergence theorem, since for all $b$, $0 \leq \int A_T E_B f_T(\{\beta_0 + T^{-1/2}b(t/T)\}_{t=1}^T) d\mu_T \leq 1$.

**Proof of Lemma 2:**

We first prove $T^{-1/2} \sum_{t=1}^T g_t(\theta_t) \Rightarrow N(0, V)$ in the unstable model by applying Corollary 2.7 of McLeish (1974) to $\{v'_t g_t(\theta_t)\}_{t=1}^T$ for an arbitrary fixed $v'_t v_g = 1$, which yields the desired result by the Cramer-Wold device. Note that $T^{-1} \sum_{t=1}^T E[|g_t(\theta_t)|^{2+\epsilon} |\mathcal{G}_{t-1}] = O_p(1)$ in the unstable model implies $T^{-1} \sum_{t=1}^T E[|g_t(\theta_t)|^{2+\epsilon}] > T^{1/2} a |\mathcal{G}_{t-1}] \overset{p}{\to} 0$ for all $0 < a < \infty$ in the unstable model. To invoke Corollary 2.7 of McLeish (1974) it thus remains to show that $T^{-1/2} \sup_{t \leq T} ||g_t(\theta_t)|| \overset{p}{\to} 0$ and $||T^{-1} \sum_{t=1}^T g_t(\theta_t) g_t(\theta_t)' - V|| \overset{p}{\to} 0$ 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 $\theta$ on the line segment between $\theta_0$ and $\theta_t$.

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

$$T^{-1} \sup_{t \leq T} ||\tilde{G}_t|| \leq pT^{-1} \sup_{t \leq T} \sup_{\theta \in \Theta} ||G_t(\theta) - G_t(\theta_0) + G_t(\theta_0)||$$

$$\leq pT^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta} ||G_t(\theta) - G_t(\theta_0)|| + pT^{-1} \sup_{t \leq T} ||G_t(\theta_0)||.$$

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

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

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

where

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

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

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

$$T^{-1} \sum_{t=1}^{T} ||\tilde{G}_t|| \leq pT^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta} ||G_t(\theta) - G_t(\theta_0) + G_t(\theta_0)||$$

$$\leq pT^{-1} \sum_{t=1}^{T} \sup_{\theta \in \Theta} ||G_t(\theta) - G_t(\theta_0)|| + pT^{-1} \sum_{t=1}^{T} ||G_t(\theta_0)||$$

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

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

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

**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 \(\beta_0\) and \(\beta_t\), so that by the same arguments used in the proofs of Lemmas 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 \(z_t = v_0 B(t/T) s_t(\beta_0) + v_1' s_t(\beta_0)\), \(\{z_t, \tilde{\xi}_t\}\) is a m.d. array with conditional variance

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

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

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

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

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

and the result follows.
References


Table 1: Rejection Probabilities of 5% Nominal Tests, $T = 100$ and $r = 0.5$

| Break     | $h$ | $t_1$ | $t_2$ | $N_{\text{all}}$ | $N_1$ | $N_2$ | $M_1$ | $M_2$ | Random walk | $t_1$ | $t_2$ | $N_{\text{all}}$ | $N_1$ | $N_2$ | $M_1$ | $M_2$
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho_{XZ} = 0$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0</td>
<td>6.4</td>
<td>6.6</td>
<td>4.3</td>
<td>4.7</td>
<td>4.6</td>
<td>4.6</td>
<td>4.6</td>
<td>4.6</td>
<td>6.4</td>
<td>6.6</td>
<td>4.3</td>
<td>4.7</td>
<td>4.6</td>
<td>4.6</td>
<td>4.6</td>
<td>4.6</td>
</tr>
<tr>
<td>sm</td>
<td>6.8</td>
<td>6.9</td>
<td>32.1</td>
<td>50.2</td>
<td>4.4</td>
<td>49.8</td>
<td>4.4</td>
<td>6.9</td>
<td>7.3</td>
<td>33.6</td>
<td>43.1</td>
<td>4.8</td>
<td>43.0</td>
<td>4.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lg</td>
<td>7.6</td>
<td>7.9</td>
<td>87.0</td>
<td>96.1</td>
<td>4.1</td>
<td>96.1</td>
<td>3.7</td>
<td>7.4</td>
<td>8.4</td>
<td>59.5</td>
<td>67.5</td>
<td>4.8</td>
<td>68.0</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{XZ} = 0.5$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sm</td>
<td>6.7</td>
<td>6.9</td>
<td>32.1</td>
<td>50.2</td>
<td>14.3</td>
<td>39.3</td>
<td>4.4</td>
<td>6.8</td>
<td>7.3</td>
<td>33.6</td>
<td>43.1</td>
<td>14.5</td>
<td>36.7</td>
<td>4.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lg</td>
<td>7.6</td>
<td>7.9</td>
<td>87.0</td>
<td>96.1</td>
<td>30.7</td>
<td>89.7</td>
<td>3.7</td>
<td>7.7</td>
<td>8.4</td>
<td>59.5</td>
<td>67.5</td>
<td>23.1</td>
<td>62.6</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho_{XZ} = 0.9$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>sm</td>
<td>6.9</td>
<td>6.9</td>
<td>32.1</td>
<td>50.2</td>
<td>41.1</td>
<td>12.9</td>
<td>4.4</td>
<td>7.1</td>
<td>7.3</td>
<td>33.6</td>
<td>43.1</td>
<td>36.9</td>
<td>15.0</td>
<td>4.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lg</td>
<td>7.9</td>
<td>7.9</td>
<td>87.0</td>
<td>96.1</td>
<td>87.7</td>
<td>35.0</td>
<td>3.7</td>
<td>8.3</td>
<td>8.4</td>
<td>59.5</td>
<td>67.5</td>
<td>60.0</td>
<td>32.1</td>
<td>5.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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.

Figure 2: Small Sample Rejection Probabilities of 5% Nominal Tests in the New Keynesian Phillips Curve Data Generating Process as Function of the Magnitude of the Instability.