The Econometrics of the Lucas Critique:
Estimation and Testing of Euler Equation Models with Time-varying Reduced-form Coefficients

Hong Li *†
Princeton University

Abstract

The Lucas (1976) critique argued that the parameters of the traditional unrestricted macroeconometric models were unlikely to remain invariant in a changing economic environment. Since then, rational expectations models have become a fixture in macroeconomics − Euler equations in particular. An important, though little acknowledged, implication of the Lucas critique is that testing stability across regimes should be a natural diagnostic for the reliability of Euler equations. This paper formalizes this assertion econometrically in the framework of the classical two-step minimum distance method: The time-varying reduced-form in the first step reflects private agents' adaptation of their forecasts and behavior to the changing environment; The presumed ability of Euler conditions to deliver stable parameters that index tastes and technology is interpreted as a time-invariant second-step model. Within this framework, I am able to show, complementary to and independent of one another, *both* standard specification tests *and* stability tests are required for the evaluation of Euler equations. Moreover, this conclusion is shown to extend to other major estimation methods. Following this result, three standard investment Euler equations are submitted to examination. The empirical results tend to suggest that the standard models have not, thus far, been a success, at least for aggregate investment.

JEL Classification: C13; C52; E22.

Keywords: Parameter instability; Model validation testing; Classical minimum distance estimation; The Lucas critique; Euler equations; Aggregate investment.

---

* Department of Economics, Princeton University, NJ 08544. E-mail: hli@princeton.edu. First draft: August 2004. This version: October 2004.

† I would like to thank Mark Watson for numerous discussions that lead to the idea of this paper, and for advice and encouragement throughout the project. I am also grateful to Ulrich Mueller for many valuable discussions. Part of the theoretical work in this paper is based on the joint research in Li and Mueller (2004). Many thanks to Bo Honore for continuous encouragement and helpful comments. My thanks also extend to Gregory Chow and Mike Woodford for useful suggestions. All mistakes are mine.

Table of Contents

1 Introduction

The Lucas critique of econometric policy evaluation argues that it is inappropriate to estimate econometric models of the economy, in which endogenous variables appear as unrestricted functions of exogenous or predetermined variables, if one proposes to use such models for the purpose of evaluating alternative economic policies. Since then, forward-looking models have become a fixture in macroeconomics literature. A wide variety of dynamic stochastic theories give rise to linearized Euler conditions of the generic form

$$\alpha(L)x_t \;=\; E_t\left(\psi(L)z_t\right) + \rho(L)s_t + e_t \tag{1}$$

where x_t is the decision variable, z_t and s_t are other variables of the system where z_t often includes x_t, and e_t is an unobservable disturbance term, $\alpha(L)$, $\psi(L)$ and $\rho(L)$ are polynomial lag operators, and $\psi(L)$ is two-sided. E_t denotes rational expectations conditional on date t information set. The popularity of such models derives from the fact that they make the notion of forward-lookingness in economic decision explicit and address the Lucas (1976) critique.

The estimation of structural parameters and testing for validity of Euler equations have been subjects of considerable research. To deal with the common feature of these models that one variable often depends on the contemporaneous and expected future values of endogenous variables, various approaches have been proposed in the literature. In this paper, I focus on the classical minimum distance estimation method, which explores the cross-equation restrictions imposed by a structural equation on the reduced-form model. This is a standard method in econometrics literature[1] and cross-equation restrictions have been widely used in macroeconomic applications[2]. Let the reduced-form model take the form of

$$y_t \;=\; w_t'\phi + \varepsilon_t \tag{2}$$

where y_t is a vector of endogenous variables including x_t, z_t and s_t in (1) and ϕ is a vector of reduced-form coefficients. A hallmark of any structural rational expectations model is that it imposes restrictions on the reduced-form coefficients. Intuitively, the description of x_t in the structural model should be consistent with the description of x_t in the reduced-form model.

Let θ denote the vector of structural parameters of interest[3]. Derivation of the cross-equation restrictions is a straightforward application of Campbell-Shiller (1987) methodology[4], which results in

$$g(\phi,\ \theta) \;=\; 0$$

[1]See Chiang (1956), Ferguson (1958), Newey and McFadden (1994) and Hayashi (2000).

[2]See Fuhrer and Moore (1995), Jondeau and Lehihan (2001), King and Kurmann (2002), Kurmann (2003), Sbordone (2003) and Li (2004), just to name a few.

[3]For example, in the context of equation (1), $\theta = [\{\alpha_i\}, \{\psi_i\}, \{\rho_i\}]$.

[4]In addition to Campbell and Shiller (1987), Sargent (1979) also points out that Euler equations of dynamic stochastic theories impose cross-equation restrictions on the coefficients of reduced-form forecasting processes.

that links the reduced-form coefficients and the structural coefficients. Assuming a stationary environment, θ can be estimated using the standard two-step procedure: the reduced-form model is estimated in the first step to obtain an estimator $\widehat{\phi}$, the structural parameters in θ are solved in the second step using $g(\widehat{\phi},\ \theta)$ as the moment condition.

However, the specification of the reduced-form model (2) explicitly assumes that the reduced-form coefficients are time-invariant. A growing number of empirical studies have looked at the instability of reduced-form empirical models. Stock and Watson (1996) provides the most general evidence by investigating the stability of a large set of macroeconomic variables. They find wide-spread instability in both univariate and bivariate models[5]. Moreover, the authors point out, the results obtained from low-dimensional models are not restrictive because the instability in the low-dimensional systems implies the instability in larger systems[6]. Therefore, empirical evidence suggests that a time-varying reduced-form model $y_t = w_t'\phi_t + \varepsilon_t$ is more appropriate in many cases. In the two-step minimum distance problem, this leads to cross-equation restrictions of the form

$$g(\phi_t,\ \theta) \ = \ 0.$$

This paper studies the case in which ϕ_t varies in a *persistent* manner[7]. To capture the persistence, the time-varying parameters (TVP) model is considered,

$$\phi_t = \phi_{t-1} \ + \ v_t.$$

Why do I consider the TVP model? The underlying changes in the structure of the economy could affect the parameters in different ways. Some might be of a discrete nature whereas others might be of a more gradual nature. However, when the true form of the instability is unknown, the flexibility of the TVP specification makes it an appealing model because it can approximate discrete changes and many other types of instabilities[8]. Hence, the TVP model provides a parsimonious way of capturing a general form of instability.

[5] Similar results have been obtained in empirical finance. Lettau and Ludvigson (2001) and Paye and Timmermann (2004) find instability in return forecasting model.

[6] In the context of vector autoregressions, this is confirmed by Boivin (1999) which concludes that there is compelling evidence of instability in monetary vector autoregressions.

[7] Temporary time variations are not discussed because they do not necessarily cause the same difficulties in estimation and inference as the persistent form of time variation

[8] Regarding its flexibility, Nyblom (1989) points out that, the TVP specification $\phi_t = \phi_{t-1} + v_t$ can not only describe a gradual change in ϕ_t when the disturbance term v_t follows a continuous distribution, it may also model discrete breaks that occur at randomly chosen time points during the observation period. Most recently, Elliot and Muller (2003) provide theoretical support for this argument by showing that tests for a very general set of breaking models (including the TVP model, frequently occurring breaks, clustered breaks, etc.), are asymptotically equivalent. The ability of the TVP model to uncover discrete breaks and other forms of instabilities can be illustrated by simulation. In addition to its flexibility, a second advantage of the TVP model is that the number of the parameters is small and independent of the true number of breaks.

Empirical studies, such as Stock and Watson (1996), also find that although a substantial fraction of macroeconomic relationships are unstable, in most cases this instability is characterized by *small period-to-period* variations in the coefficients[9] In the TVP setup, this empirical feature can be formally modeled as

$$v_t = \tau \nu_t \quad \text{where} \quad \tau = \lambda / T$$

where τ and λ are scalars governing the size of the variance of the parameters. Briefly speaking, by making the variance of the parameters local-to-zero, this parameterization provides a meaning to it being "small period-to-period". Moreover, the nesting of λ/T is a useful device in obtaining local asymptotic results. (See Section 2.2 for more discussion.) For these reasons, this modeling strategy has been becoming a standard treatment in modeling small but persistent instability[10].

In short, motivated by empirical findings, this paper deals with the estimation and validation of a structural relation with *constant* coefficients[11], while the reduced-form relation is an unstable one with coefficients that are *time-varying* in a small but persistent manner.[12].

From an *economic* viewpoint, the sort of cross-equation restrictions studied in the paper reflects the basic thrust of the Lucas (1976) critique. On the one hand, the reduced-form coefficients, ϕ, reflect private agents' understanding of the underlying economic environment. As argued in Lucas (1976) paper, ϕ would be functions of more fundamental structural parameters and parameters that describe the characteristics of economic policy rules. Accordingly, when there were any structural change in the laws of motion for the variables, say, induced by any change in policy regime, the reduced-form model would generate forecasts of future values of variables which vary with policy changes, if the reduced-form model correctly approximate the way in which agents form expectations. So, the coefficients of the reduced-form model

[9]For instance, in Stock and Watson's (1996) investigation of 76 U.S. macroeconomic series, the outcomes of the stability tests are often found to be borderline significant. Note that instabilities that are relatively small compared to the sample information will be *too* difficult to be detected and should not matter for inferences based on asymptotic theory. On the other hand, instabilities that are large relative to sample information should be detected by statistical tests *for sure*. Obviously the empirical evidence obtained in Stock and Watson's (1996) comprehensive study does not support the two cases. Rather, it points to an intermediate case.

[10]See for example, Stock and Watson (1996), Stock and Watson (1998), Boivin (1999) and Elliott and Muller (2004), among others. It is worthwhile to point out, the modeling strategy of a random walk with local-to-zero device was used in the above mentioned work for stability testing. In the current paper, the same modeling strategy is employed for the purpose of estimation and inference.

[11]That is, I implicitly assume that the instability in the reduced-form model is induced by changes in the underlying structure of the economy *other than* the structural equation under study. In this sense, the estimation procedure developed in this paper is not applicable to unstable structural relations. For the estimation procedure designed for an unstable rational expectations structural relations, see for example, Boivin and Watson (1999).

[12]In the context of a stable structural model, such as Euler equation (1), the instability enters the cross-equation restrictions through the *unstable forecasts* of future values of the variables produced by the time-varying forecasting model.

could not be expected to remain stable across time and regimes. On the other hand, the genesis of the Euler equation approach is exactly because the reduced-form parameters are unstable. A correctly specified Euler condition is supposed to possess the ability to deliver stable deep parameters of tastes and technology. Under this null hypothesis, the vector of structural coefficients in the cross-equation restriction, θ, which index private agents' optimality condition, would be expected to be stable over time.

From an *econometric* viewpoint, the modification from a constant ϕ to a drifting ϕ_t transforms the econometric model from a standard textbook problem to a two-step TVP problem that has not been thoroughly understood in the literature. Consequences could be severe if the time variation is ignored. First, standard estimation, inference and model evaluation of the second-step may be invalidated through the first-step estimation, which is distorted by the ignored time variation. The second-step estimator and test statistics may cease to follow the standard distribution, even asymptotically, and hence lead to misleading results. Second, if regressors in a reduced-form model contain lagged dependent variables, as in the widely-used autoregression or vector autoregression model, the unit root contained in ϕ_t would unavoidably induce non-stationary regressors in the reduced-form model[13], which would further complicate the second-step estimation and testing. It follows from the above discussion that, a desired estimation and procedure should appropriately take the time variation into account.

In the literature, the traditional estimation procedure for TVP models is by maximum likelihood estimation, making use of Kalman filter to construct the likelihood function. Suppose the amount of instability in ϕ_t is large enough relative to the sample information, the traditional ML procedure for TVP models can be applied to estimate θ in the two-step minimum distance problem[14]. If, on the other hand, the variance of the parameters are small, which, as I have argued, is indeed the empirically relevant case, the MLE will encounter numerical problem and break down[15].

The inference problems of ignoring the time variations when it is present, together with

[13] To see this problem, consider an unstable AR(1) model: $y_t = \phi_t y_{t-1} + \varepsilon_t$ with ϕ_t following (9) and (10). The feed back from ϕ_t to future values of the regressors is obvious. To see the consequence of such a feedback, it is straightforward to solve y_t, as a function of ε_t and ϕ_t, $y_t = \varepsilon_t + \sum_{j=1}^{\infty} \left(\prod_{i=1}^{j-1} \phi_{t-i} \right) \varepsilon_{t-j}$. Thus, even if ε have the desired stationarity and independence, the distribution of the regressor, the lagged values of y_t, is non-stationary in a complicated manner.

[14] This can be done in two steps: (1) write the constrained regression $y_t = w_t'\phi_t + \varepsilon_t$ s.t. $g(\phi_t, \theta_0) = 0$ as an unconstrained regression $y_t = x_t'\theta_0 + z_t'\gamma_t + \varepsilon_t$. by some proper transformation. (2) impose the TVP specification and normality, implement MLE through Kalman Filtering. Note that in the first step, $g(\phi_t, \theta_0) = 0$ is typically a non-linear restriction, and as a result, θ_0 would be non-linear in ϕ_t, i.e., $\theta_0 = \theta(\phi_t)$. But it is possible to show, $\theta(\phi_t) = M\phi_t + O_p(T^{-1})$ where M is a constant matrix. Thus the non-linearity in $\theta(\phi_t)$ as a function of ϕ_t is negligible.

[15] The MLE tends to estimate the variance to be zero. This is because the MLE has a large point mass at zero, which is related to the "pile-up" problem in the literature. See Stock and Watson (1998) for a more detailed discussion on this issue.

the limitations of the traditional MLE procedure in handling empirically relevant instabilities, provide the econometric motivation of the present paper. But, rather than directly handling ϕ_t, as in the MLE procedure, this paper takes a different track. Note that the parameters of interest, θ, in the two-step problem are constant (according to the underlying economic theory), and the instability is solely from ϕ_t. This feature naturally prompts me to consider whether the standard procedure (assuming erroneously a constant ϕ) is possible to remain valid in the presence of the time variation in the first step model.

Looking ahead to my results, the *econometric* findings can be summarized as follows. First, in the context of the two-step method, I find that:

- Regarding the estimation and inference of the structural (i.e., the second-step) equation, nothing is lost by ignoring the reduced-form (i.e., the first-step) instability. The standard minimum distance method remains valid for the second-step estimation, even though the first-step estimation is unavoidably contaminated. Accordingly, the standard treatment leads to valid inference about the coefficients in the structural equation.
- The conventional specification test for the structural (i.e., the second-step) model, ignoring the reduced-form (i.e., the first-step) instability, remains valid. However, the conventional test alone is not enough for an overall model evaluation − it has no power in detecting a class of local alternatives. Rather, an overall validation of the second-step structural model consists *not only* of the standard test for the cross-equation restrictions, *but also* of the assessment of an additional stability requirement.
- To assess the stability requirement of the the second-step structural model, it is shown that a desired test should be able to isolate the instability in the coefficients *being tested* from the instability induced by the *non-tested* coefficients. I verify that many existing stability tests possess this property and can be applied unaltered.
- All the above results apply to the case in which the reduced-form model is an unstable autoregression or an unstable vector autoregression. Recall a first-step model of this kind has non-stationary regressors induced by the persistence in the time-varying coefficients. The analysis conducted in the paper concludes that such non-stationarity in the first-step regressors *does not* have any asymptotic effects, at least for the kind of instabilities this paper is concerned with, and hence can be safely ignored.

Based on these findings, I propose a relatively simple solution to the seemingly complex problem of estimating and testing the two-step model with small reduced-form time variation: the standard textbook procedure plus the assessment of a stability requirement for which a battery of well-developed tools are available.

Second, I show that the above results, especially those on model validation testing, are not limited to the two-step minimum distance method, and are not limited to linear frameworks. Rather, they are general results that extend to many other methods, such as the generalized method of moments and the maximum likelihood method. As a methodological note, my analysis concludes that, regardless of the estimation approach, when a restriction under examination involves coefficients varying in a small but persistent manner, say, in the generic form of $a(\phi_t) = 0$, it can be evaluated by sequentially conducting two tests: (i) the standard specification test for the stable restriction $a(\phi_0) = 0$, ignoring the instability in the coefficients; (ii) a stability assessment testing for the stability requirement of the restriction. Moreover, the test in (i) and the test in (ii) are shown to be independent asymptotically, which makes it easy to have an overall test with a designed significance level.

Regarding contribution of the paper from an *economics* point of view, I note that although apparently little received in empirical research, the assertion that, testing the stability over time or across regimes should be one important diagnostic for the reliability of macroeconomic models, especially Euler equations, has been acknowledged in previous literature, though mainly on the basis of the economic rationale of the Lucas (1976) critique, see Ghysels and Hall (1990) and Oliner, Rudebusch and Sichel (1995). In the present paper, this assertion is formalized econometrically, for the two-step minimum distance method and is shown to extend to other major econometric approaches. As a result, my analysis strongly recommend a re-orientation of the evaluation of macroeconomic models: together with standard specification tests, stability tests should be routinely reported[16].

Finally, in the application part of the paper, using the two-step minimum distance method, I examine the empirical relevance of several investment Euler equations that are typical of those used in the literature. In my application, reduced-form vector autoregressions are used as the first-step models, to summarize the dynamics of the economy. Persistent and small instability is found in the vector autoregressions, which justifies the estimation and model validation procedure I have developed. Although Euler equations have been a very popular approach to modeling investment and there have been very little skepticism about these standard investment models, my empirical results tend to suggest that the standard models have not, thus far, been a success, at least for aggregate investment. In particular, one of the Euler equations under examination, which is most widely-used in the literature, exhibits a considerable degree of parameter instability.

The rest of the paper is organized as follows. Section 2 describes the two-step model with time-varying parameters in the first step. Technical assumptions required to obtain asymptotic

[16]There is a strand of research in the literature concerned with testing the Lucas critique, based on the concept of *super exogeneity*. See for example, Hendry and Richard (1983), Engle, Hendry and Richard (1983), Hendry (1988), Hendry and Ericsson (1991a and 1991b), Ericsson (1992), Baba, Hendry and Starr (1992), Favero and Hendry (1992), Engle and Hendry (1993) and Ericsson and Irons (1994), among others. This literature focuses on testing for the stability of the reduced-form coefficients, i.e., ϕ_t in the reduced-form model $y_t = w_t'\phi_t + \varepsilon_t$, provided that ϕ_t is a function of the underlying unstable coefficients in the policy equation ($-$the source of the potential reduced-form instability according to the Lucas critique). Different from this literature, the starting point of the current paper is that the reduced-form coefficients, ϕ_t, are unstable and the focus is on the structural equations that by construction are supposed to be stable across time and regimes.

results are discussed. Section 3 deals with the estimation of the two-step time-varying parameters model. Section 4 studies model validation testing of the two-step time-varying parameters model. Section 5 generalizes the results obtained in preceding sections to other econometric approaches. Section 6 presents the application to investment Euler equations. Section 7 concludes. Mathematical details are in the appendix.

2 The Model and Assumptions

This section lays out the two-step TVP model in which the first-step model involves small but persistent time variation. Conditions needed to obtain the asymptotic results are provided and discussed. To start, I review the standard (non-TVP) two-step model and summarize the main asymptotic results.

2.1 The Standard Two-step Model

The purpose of this review section is twofold: to establish the notation for the subsequent sections and to provide a basis for future comparison. The standard two-step model is

$$\begin{aligned} \text{1st step}: \quad & y_t \;=\; w_t'\phi_0 \;+\; \varepsilon_t \\ \text{2nd step}: \quad & 0 \;=\; g(\phi_0, \theta_0) \end{aligned} \tag{3}$$

where $\{y_t, w_t\}_{t=1}^T$ are observed variables, w_t is a $p \times 1$ vector of regressors, ε_t is a mean zero random disturbance, ϕ_0 is a $p \times 1$ vector of reduced-form coefficients, θ_0 is a $r \times 1$ vector of structural coefficients and $g(\phi_0, \theta_0)$ is a $l \times 1$ vector of restrictions imposed on ϕ_0 and θ_0. θ_0 is the parameters of interest. Assuming stationarity, the two-step estimation is straightforward. In the first step, the reduced-form model is estimated by OLS. Let $\widehat{\phi}$ denote the OLS estimator of ϕ_0,

$$T^{1/2}\ (\widehat{\phi} - \phi_0) \;\Rightarrow\; \mathcal{N}\,(0,\ V_\phi) \tag{4}$$

where $V_\phi = E(w_t w_t')^{-1}\text{Var}(w_t\varepsilon_t)E(w_t w_t')^{-1}$. In the second step, θ is estimated as the solution to $\min_\theta\ Q_T(\theta) = g(\widehat{\phi}, \theta)'\ W_T\ g(\widehat{\phi}, \theta)$ where W_T is some weighting matrix. The optimal weighting matrix is determined by $g(\widehat{\phi}, \theta_0)$. A mean value expansion around $g(\phi_0, \theta_0)$ gives

$$T^{1/2}\ (g(\widehat{\phi}, \theta_0) - g(\phi_0, \theta_0)) \;=\; D_g\ T^{1/2}\ (\widehat{\phi} - \phi_0) \;+\; o_p(1) \tag{5}$$

where $D_g = \partial g(\phi_0, \theta_0)/\partial\phi$ is an $l \times p$ matrix, with rank l. Therefore, the asymptotic variance of $g(\widehat{\phi}, \theta_0)$ is computed as $\text{Avar}(g(\widehat{\phi}, \theta_0)) = D_g V_\phi D_g'$ and the optimal weighting matrix W_{opt} equals $\text{Avar}(g(\widehat{\phi}, \theta_0))^{-1}$. The resulting second-step estimator, $\widehat{\theta}$, is a function of the first-step estimator, $\widehat{\phi}$, with the limiting distribution

$$T^{1/2}\ (\widehat{\theta} - \theta_0) \;=\; M D_g\ T^{1/2}(\widehat{\phi} - \phi_0) + o_p(1) \;\Rightarrow\; \mathcal{N}\,(0,\ V_\theta) \tag{6}$$

where $M = [G_\theta(\phi_0, \theta_0)' W_{opt} G_\theta(\phi_0, \theta_0)]^{-1} G_\theta(\phi_0, \theta_0)' W_{opt}$ with $G_\theta(\phi, \theta) = \partial g(\phi, \theta)/\partial\theta$. The efficient asymptotic variance for $\widehat{\theta}$ is $V_\theta = \left[G_\theta(\phi_0, \theta_0)'(D_g V_\phi D_g')^{-1} G_\theta(\phi_0, \theta_0)\right]^{-1}$. The J-test for the cross-equation restrictions is

$$J_T \;=\; T\; g(\widehat{\phi}, \widehat{\theta})'\; \widehat{W_{opt}}\; g(\widehat{\phi}, \widehat{\theta}) \;\Rightarrow\; \chi^2_{(l-r)}. \tag{7}$$

Essentially, all asymptotic results in the second step are built up on the asymptotics of the first-step estimator, $\widehat{\phi}$. The limiting distribution of $g(\widehat{\phi}, \theta_0)$ is the heart of the second-step estimation and inference. It determines the optimal weighting matrix, and hence the behavior of $\widehat{\theta}$ and J_T.

2.2 The Two-step Model with Time-varying Coefficients

When there is time variation in the reduced-form, together with the maintained assumption that the structural parameter θ_0 is stable, the two-step model becomes

$$\begin{aligned} \text{1st step}: \quad & y_t \;=\; w_t'\phi_t + \varepsilon_t \\ \text{2nd step}: \quad & 0 \;=\; g(\phi_t,\ \theta_0), \quad t = 0, ..., T \end{aligned} \tag{8}$$

where θ_0 is the vector of parameters of main interest. The implications of the model are simple. It allows the coefficients on the reduced-form regressors w_t to evolve over time in a non-stationary manner and hence, alters the structural of the relationship between y_t and w_t within a maintained linear form. But it does so in a way that preserves a time-invariant, and possibly non-linear relationship between the values of the reduced-form coefficients and the structural coefficients over time.

As mentioned in the introduction, motivated by the empirical evidence observed in macroeconomic relations, the type of time variation this paper is concerned with is of the form

$$\phi_t \;=\; \phi_{t-1} + v_t \qquad \text{with} \qquad v_t \;=\; \tau\nu_t \tag{9}$$

where τ is a scalar, $\nu_t \sim (0,\ \Sigma)$ with Σ being a non-zero matrix, ν_t and ε_t are serially and mutually uncorrelated, mean zero random disturbances. Several extensions could be made for the TVP specification in (9). First, in addition to the persistent time variation in ϕ_t, there may be temporary time variation in ϕ_t as well, so that ν_t may be serially correlated[17]. Second, rather than containing a unit root, ϕ_t may contain a near unit-root[18]. Since all these extensions or modifications do not change the main results of the paper in a substantive way, in what follows, for simplicity, I will continue to use the basic TVP specification in (9).

[17]In this case, ν_t can be modeled as $\nu_t = B(L)\mu_t$ where μ_t and ε_t are serially and mutually uncorrelated. As long as $B(L)$ is one-summable and $B(L) \neq 0$, asymptotic results obtained in the serially-uncorrelated-ν_t case can be easily extended to the serially-correlated-ν_t case with some small modifications.

[18]When ϕ_t contains a near unit root, it can be modeled by using the local-to-unity device, $\phi_t = (1-c/T)\phi_{t-1} + v_t$. In my problem, asymptotic results obtained in the exact unit-root case extends to the near unit-root case by replacing the Wiener process associated with ϕ_t with a corresponding diffusion process that depends on c.

Note that decomposing $\phi_t - \phi_{t-1}$ into τ and ν_t has the advantage that with Σ being some known matrix[19], the magnitude of the instability, represented by the the variance of the time-varying parameters, is governed by a scalar, τ, only. This greatly simplifies the analysis because the empirical feature of the small period-to-period variation in the *vector*, ϕ_t, can be formalized by making this scalar, τ, local-to-zero,

$$\tau = \lambda/T. \tag{10}$$

Several comments regarding (10) are in order. First, by this parameterization, although the instability vanishes asymptotically, in any finite sample, the amount of instability is small but non-zero[20]. Second, estimation and inference are based on asymptotic distributions. Nesting (10) is exactly the asymptotic device that would be used in computing the local asymptotic power of the stability tests under a TVP alternative. It makes the distribution of the tests non-degenerate and dependent on λ. Thus, the magnitude of the kind of small instabilities considered in the paper is large enough to invalidate inference that ignores them.

2.3 Conditions and Assumptions

To obtain the theoretical results, it is necessary to impose some conditions on the data generating processes. The set of conditions which will be presented is by no means the weakest possible set of sufficient conditions, although most of the conditions are fairly regular in econometrics analysis. In particular, these conditions intend to cover time-varying parameters autoregression and vector autoregression models. As mentioned in the introduction, for this class of models, the regressors contain lagged dependent variables, so instability in the coefficients may create estimation and inference problem: it will introduce feedback from the unstable coefficients to future values of regressors, and hence induce non-stationarity in the regressors. If this problem is not addressed, the scope within which the analysis of this paper is applicable would be much limited, given the extensive use of this class of models in applied macroeconomics. As detailed below, this problem is circumvented by making use of a concept called *contiguity*, which has existed in the statistical literature for nearly half a century, and is introduced to econometric analysis only recently. So long as the property of contiguity holds for the data generating process, non-stationarity in the regressors caused by the unstable coefficients is negligible asymptotically.

Since the conditions introduced in this section will be used for a range of models in the rest of the paper, it is useful to digress from the two-step minimum distance framework and to provide these technical requirements in the context of a general model which nests all specific

[19]In practice, some suitable normalizations can be imposed on Σ to make it proportional to some matrix known to the econometricians. See, Stock and Watson (1998) and Nyblom (1989) for more discussion.

[20]To see this, note that $\mathrm{Var}(\phi_t - \phi_{t-1}) = O(T^{-2})$. As T goes to infinity, $\mathrm{Var}(\phi_t - \phi_{t-1})$ approaches 0. But for a fixed sample size T, its magnitude not zero for any non-zero λ.

TVP models that will be encountered latter in the paper. Thus, consider a TVP model of the form

$$y_t \;=\; w_{1t}'\phi_{1t} + w_{2t}'\phi_{2t} + \varepsilon_t, \tag{11}$$

where w_{1t} and w_{2t} are $k_1 \times 1$ and $k_2 \times 1$ vectors of regressors, ε_t is the error term and ϕ_{1t} and ϕ_{2t} are $k_1 \times 1$ and $k_2 \times 1$ vectors of coefficients that might be time-varying. With the decomposition of $\phi_t = [\phi_{1t}' \ \phi_{2t}']'$, the TVP process for the coefficients can be rewritten as $\phi_{it} = \phi_{it-1} + v_{it}$, $v_{it} = \tau_i \nu_{it}$ and $\tau_i = \lambda_i / T$ for $i = 1$ and 2. τ_i's and λ_i's are scalars. $\mathrm{Var}(\nu_{it}) = \Sigma_i$.

Suppose ϕ_{1t} is the vector of coefficients of interest, ϕ_{2t} is the vector of nuisance coefficients. Several versions of model (11) will be discussed in the paper. The model with a constant ϕ_1 and an unstable ϕ_{2t} is discussed in Section 3.2. The model with ϕ_{1t} being possibly time-varying and ϕ_{2t} being time-varying is the testing model discussed in Section 4.2. The model in which ϕ_{1t} is possibly time-varying but ϕ_2 is constant is the testing model that has been examined in the literature. The reduced-form model in (8) is a special case of (11) with $w_{2t} = 0$.

In the rest of the paper, I use "*under* ϕ_t" to refer to the case in which the data are generated by the *true* model $y_t = w_{1t}'\phi_{1t} + w_{2t}'\phi_{2t} + \varepsilon_t$ which has time-varying coefficients. On the other hand, "*under* ϕ_0" refers to the case in which data are generated by the corresponding *hypothetical* model $y_t = w_{1t}'\phi_{10} + w_{2t}'\phi_{20} + \varepsilon_t$ where all coefficients are stable. In addition, let "$\Rightarrow$" denote weak convergence to the relevant stochastic process. Let "$\xrightarrow{p}$" denote convergence in probability. W_i's and W_{ϕ_i}'s (for $i = 1$ and 2) are standard Brownian motions. W_{ϕ_i}'s are mutually independent Brownian motions associated with the ϕ_i processes, and W_{ϕ_i}'s are independent of W_i's. Let $[\cdot]$ denote the greatest lesser integer.

Condition 1: *The processes for regressors $\{w_{1t}, w_{2t}\}$ and disturbances $\{\varepsilon_t, \nu_{1t}, \nu_{2t}\}$ are such that following requirements hold under ϕ_t for $i = 1,2$ and $j = 1,2$.*

1. $T^{-1}\sum_{t=1}^{[sT]} w_{it}w_{jt}' \xrightarrow{p} s\,\Gamma_{ij}$ *uniformly in* $s \in [0,1]$ *for some constant matrix* Γ_{ij}.
2. $T^{-1/2}\sum_{t=1}^{[sT]} w_{it}\varepsilon_t \Rightarrow \Gamma_{w\varepsilon}^{i\ 1/2}\ W_i(s)$ *where* $\Gamma_{w\varepsilon}^{i}$ *is positive definite.*
3. $T^{-1}\sum_{t=1}^{[sT]} w_{it}w_{jt}'\ T^{-1/2}\sum_{k=1}^{t} \nu_{jk} \Rightarrow \Gamma_{ij}\ \Sigma_j^{1/2}\int_0^s W_{\phi_j}(r)dr.$

Condition 1 can be shown to hold in some general settings. For example, the following set of assumptions are sufficient to guarantee Condition 1. Among them, Assumptions 1 to 3 are a modified version of the assumptions used in Stock and Watson (1998). For a stationary process w_t, let $C_{i_1\cdots i_n}(r_1,\cdots,r_{n-1})$ denote the nth joint cumulant of $w_{i_1t_1},\cdots,w_{i_nt_n}$, where $r_j = t_j - t_n$, $j = 1,\cdots,n-1$, and let $C(r_1,\cdots,r_{n-1}) = \sup_{i_1,\cdots,i_n} C_{i_1\cdots i_n}(r_1,\cdots,r_{n-1})$.

Assumption 1: The regressors $\{w_{1t}, w_{2t}\}$ are stationary *under* ϕ_0 with eighth order cumulants that satisfy $\sum_{r_1,r_2,r_3=-\infty}^{\infty} |C_1(r_1,r_2,r_3)| < \infty$ and $\sum_{r_1,r_2,r_3=-\infty}^{\infty} |C_2(r_1,r_2,r_3)| < \infty$.

This assumption requires that regressors to have bounded moments and are not integrated of order one or higher *under* ϕ_0. Note that the assumption of stationarity of the data generating process is only required in the *hypothetical* stable model, which is relatively easy to satisfy, especially when the regressors involve lagged dependent variables.

Assumption 2: $(\varepsilon_t, \nu_{1t}', \nu_{2t}')'$ is a vector of *i.i.d.* errors with mean zero. ε_t has four finite moments, $(\nu_{1t}', \nu_{2t}')'$ have eight finite moments. $\{\varepsilon_t\}$ is independent of $\{\nu_{1t}, \nu_{2t}\}$.

From this assumption, the underlying shocks that generate the stochastic ϕ_{1t} and ϕ_{2t} processes are required to be independent of other sources of randomness in the regression model. The ϕ_{1t} process and the ϕ_{2t} process are allowed to be correlated.

Assumption 3: The regressors $\{w_{1t}, w_{2t}\}$ are independent of $\{\varepsilon_t, \nu_{1t}, \nu_{2t}\}$ *under* ϕ_0; Or alternatively, $\{w_{1t}, w_{2t}\}$ are independent of $\{\nu_{1t}, \nu_{2t}\}$ *under* ϕ_0 and ε_t is independent of $\{w_{1t}, w_{1t-1}, w_{1t-2}...\}$ and $\{w_{2t}, w_{2t-1}, w_{2t-2}...\}$.

Thus, the regressors are supposed to be strictly exogenous or alternatively, part or all of the regressors to be predetermined *under* ϕ_0. In particular, for the predetermined case, Assumption 3 only requires the regressors generated by the *hypothetical* stable process to be independent of the unstable coefficients in the true TVP model[21].

Finally, let z_t denote date t observations $(y_t, w_{1t}', w_{2t}')'$. Denote the joint likelihood function of the data for a given ϕ path by $f_T(z_1, ..., z_T; \phi)$. In the *hypothetical* stable model, $\phi = \phi_0$ for all t; and in the *true* model, $\phi = \{\phi_t\}_{t=1}^T$. Thus, the density of the data *under* ϕ_0 and the unconditional density of data *under* ϕ_t are $f_T(z_1, ..., z_T; \phi_0)$ and $\int_{-\infty}^{+\infty} f_T(z_1, ..., z_T; \phi)\, dv_\phi$ respectively, where v_ϕ is a measure of ϕ.

Assumption 4: Let ϕ_t follow (9) and (10). The sequence of densities $\{\int_{-\infty}^{+\infty} f_T(z_1, ...z_T; \phi)\, dv_\phi : T \geq 1\}$ are contiguous to the sequence of densities $\{f_T(z_1, ...z_T; \phi_0) : T \geq 1\}$.

Here contiguity is introduced as a high-level assumption. It may be a slight abuse the language to introduce contiguity as an *assumption*. Rather, it should be understood as a *property* possessed by the data generating process associated with the kind of instabilities described in (9) and (10). It is shown in the paper, a set of fairly general primitive assumptions on the likelihood function are sufficient to guarantee contiguity for the kind of instabilities that are

[21] To understand the meaning of Assumption 3, consider an unstable AR(1) model $y_t = y_{t-1}\phi_t + \varepsilon_t$. Under the *true* data generating process, the regressor y_{t-1} is *not* independent of the sequence $\{\phi_t, \phi_{t-1}, ...\}$. In other words, Assumption 3 will not hold *under* ϕ_t. But, under the *hypothetical* stable process $y_t = y_{t-1}\phi_0 + \varepsilon_t$, as a function of $\{\varepsilon_{t-1}, \varepsilon_{t-2}, ...\}$, the regressor y_{t-1} is indeed independent of $\{\phi_t, \phi_{t-1}, ...\}$, as long as $\{\phi_t\}$ and $\{\varepsilon_t\}$ are mutually independent. In addition, note that Assumption 3 implies conditional homoskedasticity *under* ϕ_0.

empirically relevant. But to avoid distracting the analysis from the main theme of the paper, the primitive assumptions are given in the appendix, which are followed by the proof of contiguity. Also note that though contiguity is a requirement on the likelihood, often researchers need not to know the concrete form of the likelihood in order to deal with the problems at hand. The estimation and testing of the two-step TVP model is one such example.

Briefly speaking, the concept of contiguity describes the asymptotic closeness of two sequences of densities. If a sequence of densities is contiguous to another sequence of densities, the convergence results that apply to the latter sequence also hold for the former sequences[22]. Intuitively, this says, if the likelihood functions of two data generating processes are very close to one another in an asymptotic sense, the difference between the two processes are expected to have effects that are only negligible in large samples. Therefore, relying on contiguity, in order to establish a convergence result *under* ϕ_t, one only needs to show the convergence result *under* ϕ_0. The property of contiguity will transfer convergence in probability or distribution *under* ϕ_0, to analogous convergence *under* ϕ_t[23].

Recall in an unstable autoregression or vector autoregression models, the different behavior of the regressors generated by the true data generating process and those generated by the hypothetical stable process is induced by small instability in coefficients. Thus, as a consequence of Lemma A.2 in the appendix, (i.e., the establishment of contiguity for the data generating process *under* ϕ_t,) the regressors in an unstable autoregression or vector autoregression model behaves asymptotically similarly to the regressors in the corresponding stable autoregression or vector autoregression model.

Lemma 1: *Consider the TVP model defined in (11). Suppose assumptions 1 to 4 are satisfied, then Condition 1 holds.*

This contiguity property plays a key role in obtaining Condition 1 *under* ϕ_t. Without contiguity, assumptions 1 to 3 only, (in particular assumption 3, which would not be satisfied by the autoregression or vector autoregression models *under* ϕ_t), lead to Condition 1 *under* ϕ_0. In

[22] See van der Vaart (1998) and Pollard (2001) for an introduction to the concept. For applications of contiguity in theoretical econometrics, see Andrews and Ploberger (1994), Elliott and Mueller (2004), Muller (2004), Li and Mueller (2004).

[23] To illustrate the role of contiguity, consider two examples from the unstable AR(1) model $y_t = y_{t-1}\phi_t + \varepsilon_t$ where for simplicity, assume $\varepsilon_t \sim (0,\ \sigma^2)$.

[Example 1]: Under stationarity, $T^{-1}\sum_{t=1}^{T} y_{t-1}^2 \xrightarrow{p} \frac{\sigma^2}{1-\phi_0^2}$ can be shown to hold *under* ϕ_0 (i.e., under the hypothetical stable process $y_t = y_{t-1}\phi_0 + \varepsilon_t$). Suppose the densities of data *under* ϕ_t is contiguous to the densities of data *under* ϕ_0, the above weak convergence result will also hold *under* ϕ_t (i.e., under the true unstable process), even if the regressor $\{y_{t-1}\}$ is not stationary under the true process.

[Example 2]: Let $\widehat{\phi} = (\sum_{t=1}^{T} y_{t-1}^2)^{-1}\sum_{t=1}^{T} y_t y_{t-1}$, and let $\widehat{\sigma}^2 = (T-k)^{-1}\sum_{t=1}^{T}\widehat{\varepsilon}_t^2$ where $\widehat{\varepsilon}_t = y_t - y_{t-1}\widehat{\phi}$ is the OLS residual. Under some standard assumptions, it is easy to establish $\widehat{\sigma}^2 \xrightarrow{p} \sigma^2$ *under* ϕ_0. Then suppose contiguity holds, the standard variance estimator $\widehat{\sigma}^2$ remains consistent *under* ϕ_t.

the two-step minimum distance problem, the disturbances, ε_t and ν_t , and the regressor, w_t, are assumed to satisfy Condition 1. In addition, the second-step $g(\cdot)$ function are assumed to satisfy the following regularity condition. Let Θ be the parameter space of θ and Φ_0 be some neighborhood of ϕ_0. Let $\|\cdot\|$ be the Euclidean norm.

Condition 2: *$g(\phi,\theta)$, $\partial g(\phi,\theta)/\partial\phi$ and $\partial g(\phi,\theta)/\partial\theta$ are continuously differentiable in ϕ and θ for all $\phi \in \Phi_0$ and $\theta \in \Theta$ with $\sup_{\theta\in\Theta}\sup_{\phi\in\Phi_0} \|\partial^2 g^{(i)}(\phi,\theta)/\partial\phi\partial\phi'\| < \infty$ and $\sup_{\theta\in\Theta}\sup_{\phi\in\Phi_0} \|\partial^2 g^{(i)}(\phi,\theta)/\partial\theta\partial\theta\| < \infty$ for all $i \in [1,...,dim(g)]$*[24].

Equations (8) to (10), together with Conditions 1 and 2, complete the specification of the two-step time-varying parameters model. θ_0 is the parameter of interest. $\{\phi_0, \lambda, \Sigma\}$ that characterize the random walk process of ϕ_t are the nuisance parameters.

3 Estimation of the Two-step TVP Model

3.1 Estimating by the Standard Procedure

The cross-equation restriction $g(\phi_t, \theta_0)$ is typically non-linear in ϕ_t[25]. Consider a first-order asymptotic approximation by linearizing $g(\phi_t, \theta_0)$ around ϕ_0, the initial value of the ϕ_t sequence, which can also be interpreted as the constant component of ϕ_t[26],

$$g(\phi_t,\theta_0) \;=\; g(\phi_0,\theta_0) \;+\; D_g(\phi_t-\phi_0) \;+\; O_p(T^{-1}). \tag{12}$$

where $D_g = \partial g(\phi_0, \theta_0)/\partial\phi$ is a $l \times p$ matrix with rank l, $l < p$. Derivation of (12) is in the appendix. Note that the overall restriction $g(\phi_t, \theta_0)$ at $t = 0$ gives $g(\phi_0, \theta_0) = 0$. This in turn implies the second term on the right-hand side of (12) should be zero in large samples. Thus, in order that the cross-equation restriction $g(\phi_t, \theta_0) = 0$ is satisfied in an asymptotic sense, it

[24] For many linear macro models, for example, the linear rational expectations models, $g(\phi,\theta)$ is linear in θ so that $\partial^2 g(\phi,\theta)/\partial\theta\partial\theta' = 0$. In this case, Condition 2 only requires the second derivative with respect to ϕ to be bounded.

[25] It is not too difficult to conceive of cases in which cross-equation restrictions are non-linear in ϕ_t. For instance, in the context of an Euler equation in the form of equation (1), even if it is linear in the structural coefficients θ, so that the cross-equation restriction takes the form $g(\phi_t,\theta_0) = A(\phi_t) + B(\phi_t)\theta_0$, the functions $A(\phi_t)$ and $B(\phi_t)$ are still possibly nonlinear in ϕ_t, with the nonlinearity in $A(\phi_t)$ and $B(\phi_t)$ coming from multiple-period ahead forecasts.

[26] A similar decomposition was used in Leybourne (1993) to study the time-varying coefficient regressions in the presence of *linear* restrictions. Note that linear restrictions are a special case of (12), in which the remainder term is identically zero for all t and T.

is sufficient the following condition holds.

$$\begin{aligned}
\text{Restriction } 1: \quad & 0 = g(\phi_0,\ \theta_0) && (13)\\
\text{Restriction } 2: \quad & 0 = D_g(\phi_t - \phi_0) \quad t = 1, ...T. && (14)
\end{aligned}$$

Restriction 1 is the *stable component* of the overall restriction. Inspection of (3) and (13) reveals that Restriction 1 is the same constraint one would encounter if the reduced-form model is stable. Restriction 2 is the *TVP component* of the overall restriction. So long as the first-order approximation is accurate in large samples, the instability in $g(\phi_t, \theta_0)$ that is relevant to asymptotic inference, if any, is captured by the first-order term, $D_g(\phi_t - \phi_0)$. Thus, the null hypothesis that, as a function of unstable coefficients, $g(\phi_t, \theta_0)$ takes a constant value, zero, over time requires $D_g\phi_t$, the linear combination of ϕ_t in a particular direction, stay constant at each point in time[27].

So one possible procedure to estimate θ is to perform the standard estimation using Restriction 1. However, this does not imply the standard procedure is necessarily valid. To see this, the OLS estimator of ϕ_0 is $\widehat{\phi} = [\sum w_t w_t']^{-1} \sum w_t y_t$. (In the rest of the paper, sums are taken over the whole sample period and the integrals are taken from 0 to 1 unless stated otherwise.) Writing the data generating process as $y_t = w_t'\phi_0 + (\varepsilon_t + w_t'(\phi_t - \phi_0))$, where the neglected time variation in ϕ_t is treated as an "omitted" variable, yields

$$\begin{aligned}
T^{1/2}\ (\widehat{\phi} - \phi_0) &= A_{1T} + A_{2T}\\
A_{1T} &= \left[T^{-1} \sum w_t w_t'\right]^{-1}\ T^{-1/2} \sum w_t \varepsilon_t\\
A_{2T} &= \left[T^{-1} \sum w_t w_t'\right]^{-1}\ T^{-1/2} \sum w_t w_t'(\phi_t - \phi_0)
\end{aligned}$$

where A_{1T} represents the scaled sampling error in the corresponding stable model, and as a result, leads to the standard distribution $A_{1T} \Rightarrow \mathcal{N}(0,\ V_\phi)$, where V_ϕ is the asymptotic variance in the stable model. On the other hand, A_{2T} represents the extra sampling error associated with ϕ_t, which induced a non-trivial limit, $A_{2T} = T^{-1/2} \sum(\phi_t - \phi_0) + o_p(1) \Rightarrow \mathcal{N}(0, \frac{1}{3}\lambda^2\Sigma)$. (Derivation of the limit of A_{2T} is detailed in the appendix). Thus, the standard estimation results in

$$T^{1/2} \left(\widehat{\phi} - \phi_0\right) \Rightarrow \mathcal{N}\left[0,\ V_\phi + \frac{1}{3}\lambda^2\Sigma\right] \tag{15}$$

which is a non-standard distribution[28]. From (15), it is evident the standard first-step inference of ϕ that erroneously uses V_ϕ as the asymptotic variance is misleading. The result in (15)

[27] The second restriction is true if and only if $D_g \Sigma D_g'$ where Σ is defined in (9). This in turn implies Σ is a matrix with a *reduced* rank, $dim(\phi) - dim(g)$, under the null hypothesis.

[28] The result in (15) makes use of the independence assumption between ε_t and ϕ_t, so that the asymptotic variance of $\widehat{\phi}$ is the sum of the standard variance in the stable model and the variance of the extra term induced by ϕ_t.

demonstrates that the kind of small instabilities under consideration are *not* negligible in a statistical sense since they are large enough to invalidate inference.

It would then be natural to expect the distortion extend to the second step, because the extra sampling error in A_{2T} would carry over to the second step through $g(\widehat{\phi}, \theta_0)$,

$$T^{-1/2} \ \Big(g(\widehat{\phi}, \theta_0) - g(\phi_0, \theta_0)\Big) = D_g A_{1T} \ + \ T^{-1/2} \sum D_g(\phi_t - \phi_0) + o_p(1)$$

where the effect of ϕ_t is captured by the term $T^{-1/2} \sum D_g(\phi_t - \phi_0)$. But then under the null hypothesis in (14), $D_g(\phi_t - \phi_0) = 0$ for all t. So the term representing the effect of ϕ_t disappears, and the asymptotic distribution of $g(\widehat{\phi}, \theta_0)$ is *unaffected* in the presence of ϕ_t,

$$T^{-1/2} \ (g(\widehat{\phi}, \theta_0) - g(\phi_0, \theta_0)) \ \Rightarrow \ D_g \ \mathcal{N}\,(0, \ V_\phi)\,.$$

This in turn implies, as a functional of $g(\widehat{\phi}, \theta_0)$, the asymptotic behavior of the second-step estimator, $\widehat{\theta}$, is unaffected, and the standard second-step inference remains valid. Let V_θ be the asymptotic variance of $\widehat{\theta}$ in the stable two-step model. Then the following results holds.

Proposition 1: *Consider the two-step problem described by (8)-(10). Suppose Conditions 1 and 2 holds. Under the null hypothesis of* $g(\phi_t, \theta_0) = 0$*, the standard second-step estimation is asymptotically independent of the time variation in the reduced-form model. The limiting distribution of the standard second-step estimator is given by* $T^{-1/2} \ (\widehat{\theta} - \theta_0) \Rightarrow \ \mathcal{N}\,(0, \ V_\theta)$.

3.2 Equivalent Linear TVP Representation

In this section, the first-step model is transformed in such a way that $D_g\phi_t$, the linear combination of ϕ_t of interest in (14), is isolated from the rest of the regression. The purpose of this exercise is twofold: first, to understand the insight of Proposition 1 in a linear regression setup; second, to introduce the testing model studied in the next section.

Construct a $p \times p$ matrix S by augmenting the $l \times p$ matrix D_g with a $(p-l) \times p$ matrix $H^\dagger$ such that S^{-1} exists, $S = \begin{bmatrix} D_g \\ H^\dagger \end{bmatrix}$ and $S^{-1} = [H_x' \ H_z']$ where H_x' corresponds to the first l columns of S^{-1} and H_z' corresponds to the last $(p-l)$ columns of S^{-1}. Then the first-step regression can be rewritten as

$$y_t \ = \ w_t' S^{-1} S \phi_t + \varepsilon_t \ = \ (H_x w_t)'(D_g \phi_t) \ + \ (H_z w_t)'(H^\dagger \phi_t) \ + \ \varepsilon_t.$$

Let $H_x w_t$, $H_z w_t$, $D_g \phi_t$ and $H^\dagger \phi_t$ be denoted by x_t, z_t, β_t and γ_t respectively, then the original two-step model in (8), a reduced-form model subject to the cross-equation restrictions, is equivalent to the unconstrained problem

$$y_t \ = \ x_t' \beta_t \ + \ z_t' \gamma_t \ + \ \varepsilon_t, \tag{16}$$

where β_t, if time-varying, follows $\beta_t - \beta_{t-1} = \tau_\beta \nu_{\beta t}$ with $\tau_\beta = \lambda_\beta / T$, and γ_t follows $\gamma_t - \gamma_{t-1} = \tau_\gamma \nu_{\gamma t}$ with $\tau_\gamma = \lambda_\gamma / T$, where $\text{Var}(\nu_{it}) = \Sigma_i$ for $i = \beta$ and γ. Since x_t and z_t are correlated, I specify the following partial regression

$$x_t \;=\; \vartheta' z_t \;+\; u_t. \tag{17}$$

under ϕ_0[29], where ϑ is a constant coefficient matrix and u_t is a vector of mean zero disturbances. By construction, z_t and u_t are uncorrelated, so that $E(u_t z_t') = 0$.

Table 1: Correspondences between OLS and MDE

	OLS	MDE
1.	$\beta_t = \beta_0$	$D_g\phi_t = D_g\phi_0$
2.	$\widehat{\beta}$	$g(\widehat{\phi}, \theta_0) = D_g\widehat{\phi} + o_p(T^{-1/2})$
3.	$T^{-1/2}(\widehat{\beta} - \beta_0)$	$T^{-1/2}(g(\widehat{\phi}, \theta_0) - g(\phi_0, \theta_0))$

By the transformation made above, a set of correspondences between the two-step minimum distance method (MDE) with time-varying coefficients and ordinary least square (OLS) with time-varying coefficients are summarized in Table 1. Note that under the null hypothesis of $\beta_t = \beta_0$, regression (16) becomes

$$y_t \;=\; x_t'\beta_0 \;+\; z_t'\gamma_t \;+\; \varepsilon_t \tag{18}$$

Then, according to the third correspondence in Table 1, the question investigated in the preceding subsection that, *under* $D_g\phi_t = D_g\phi_0$*, whether the distribution of* $g(\widehat{\phi}, \theta_0)$ *is affected by* ϕ_t, is equivalent to the following question: *In the linear TVP model (18) where* β *is constant, whether the distribution of* $\widehat{\beta}$ *is affected by* γ_t.

This issue is studied in Elliott and Mueller (2003b) and Li and Mueller (2004). The conclusion they obtain is, valid inference can still be made on the stable coefficients in the presence of small instability in the nuisance coefficients[30]. Therefore, the finding in Proposition 1 is not a surprise, given that the behavior of $g(\widehat{\phi}, \theta_0)$ in the two-step model parallels that of $\widehat{\beta}$ in the regression model. In fact, the result in Elliott and Mueller (2003b) and Li and Mueller (2004) extends to a more general situation where β itself is unstable. as stated in the following proposition[31]. Let V_β denote $E(u_t u_t')^{-1}\mathrm{Var}(u_t\varepsilon_t)E(u_t u_t')^{-1}$, the asymptotic variance of $\widehat{\beta}$ in the stable model.

[29]Since the *true* data generating process under consideration is a time-varying parameter model, the *true* partial regression of x_t on z_t should also be a time-varying parameter model. However, under the contiguity argument, to get the asymptotic results *under* ϕ_t, I only need to specify the relationship between x_t and z_t in the *hypothetical* stable setup.

[30]See Section 5.1 for more discussion about this finding in these papers.

[31]In the next section, this generalized independence property plays an important role in constructing appropriate tests for the constancy of β in model (16).

Proposition 2: *Let y_t obey (16)–(17). Suppose Condition 1 holds. Then, the asymptotic distribution of the OLS estimator $\widehat{\beta}$ is independent of γ_t. Its limiting distribution is given by* $T^{-1/2}(\widehat{\beta}-\beta_0) \;\Rightarrow\; \mathcal{N}(0,\; V_\beta + \frac{1}{3}\lambda_\beta^2\Sigma_\beta)$.

Several comments are in order. First, a special case of Proposition 2 is when β is stable (so that λ_β is zero). Then the inference over β is dominated by the standard term only. This is exactly the result obtained by Elliott and Mueller (2003b) and Li and Mueller (2004).

Second, in the more general case that β is unstable, the non-standard distribution of $\widehat{\beta}$ is induced by the instability in β_t itself, rather than the instability in the nuisance coefficients, γ_t. To see the intuition, when both β and γ are unstable, the standard procedure estimates

$$y_t \;=\; x_t'\beta_0 + z_t'\gamma_0 + (\varepsilon_t + x_t'(\beta_t-\beta_0) + z_t'(\gamma_t-\gamma_0)). \tag{19}$$

The instability ignored by the standard procedure are essentially two omitted variables, $x_t'(\beta_t - \beta_0)$ and $z_t'(\gamma_t-\gamma_0)$, in the composite error. I show in the appendix that whenever γ_0, the *constant* component of the nuisance coefficients, is *partialed out* in the standard procedure (by premultiplying the matrix form of equation (19) by the residual matrix $M_z = I - Z(Z'Z)^{-1}Z'$), an important by-product of this operation is, the *unstable* component of the nuisance coefficients, $(\gamma_t-\gamma_0)$, is partialed out asymptotically. From the derivation detailed in the appendix, it is clear the crucial condition behind this outcome is $E(u_t z_t') = 0$. Note that this is the orthogonal condition in the partial regression of x_t on z_t in (17), and it holds *by construction*. As a result, time variation in γ does *not* alter the limiting behavior of $\widehat{\beta}$.

4 Testing the Two-step TVP Model

With the decomposition of the overall restriction $g(\phi_t,\theta_0)=0$ into $g(\phi_0,\theta_0)=0$ and $D_g(\phi_t-\phi_0)=0$, the overall evaluation of the two-step TVP model becomes a joint test of $g(\phi_0,\theta_0)=0$ and $D_g(\phi_t-\phi_0)=0$. In this section, three issues are studied: testing for the stable restriction $g(\phi_0,\theta_0)=0$; testing for the TVP restriction $D_g(\phi_t-\phi_0)=0$; and finally, the statistical relationship between the two tests.

4.1 Testing the stable Restriction $g(\phi_0,\theta_0)=0$

The J-statistic is shown to be a functional of $g(\widehat{\phi},\theta_0)$, see the appendix. According to the analysis in Section 3.1, the asymptotic behavior of $g(\widehat{\phi},\theta_0)$ is not affected by the time variation in ϕ_t under the null hypothesis. Hence, to test for $g(\phi_0,\theta_0)=0$, the two-step TVP model can be treated as a standard two-step model because ignoring the reduced-form instability does not alter the asymptotic distribution the J-test.

Proposition 3: *Consider the two-step problem described by (8)-(10). Suppose Conditions 1 and 2 hold. Under the null hypothesis of $g(\phi_t, \theta_0) = 0$, the standard test for restriction $g(\phi_0, \theta_0) = 0$ is asymptotically independent of the time variation in ϕ_t. The asymptotic distribution of the test is given by $J_T \Rightarrow \chi^2_{dim(g)-dim(\theta)}$, with J_T being defined in (7).*

4.2 Testing the TVP Restriction $D_g(\phi_t - \phi_0) = 0$

Testing the constancy of $D_g\phi_t$ in the two-step model is essentially testing the constancy of β in model $y_t = x_t'\beta_t + z_t'\gamma_t + \varepsilon_t$, see the first correspondence in Table 1. Although there is a long-standing interest and effect in developing tests for structural stability, the studies most relevant to the current problem are concerned with testing model $y_t = x_t'\beta_t + z_t'\gamma_0 + \varepsilon_t$, where the non-tested coefficient, γ, is stable. A natural question to investigate is then, whether tests developed for *constant* nuisance coefficients are robust in the presence of *unstable* nuisance coefficients. To start, I first review the existing tests for the constancy of β when the non-tested coefficient, γ, is stable.

4.2.1 Related Literature

In the literature, a large number of stability tests have been developed for $y_t = x_t'\beta_t + z_t'\gamma_0 + \varepsilon_t$. The test focuses on the possibility of a non-constant β while requiring for other coefficients to be stable over time.

The first set of tests for this TVP model are Nyblom's (1989) locally most powerful tests and point optimal invariant tests (Saikkonen and Luukkonen (1993), Shively (1988) and Elliott and Muller (2003a)), which are invariant to translations and scale transformations. The Nyblom test is a locally optimal test obtained by comparing the Gaussian likelihood of certain maximal invariant under the null of a constant β and under the local alternative that β follows a *random walk*. Point optimal invariant tests (POI) are also Gaussian likelihood based, with one value of λ being chosen to construct the test statistic. The POI tests are optimal for the chosen value of λ.

Another set of test statistics consist of sequential Chow statistics, which are motivated to test the null of a constant β against the alternative of a single *discrete break* in β at certain fraction through the sample. Three such tests are frequently used, namely, Quandt (1960) likelihood ratio statistic (QLR), Andrews-Ploberger (1994) mean Wald statistic (MW) and Andrews-Ploberger (1994) average exponential Wald statistic (EW). These tests are performed over restricted sample period with appropriate trimming at both ends. The QLR test uses the maximum Chow statistic over all possible break dates within the trimmed sample. The MW and EW tests are optimal tests in the sense of maximizing certain weighted average power criterion over the trimmed period[32].

[32] The choice of the weighting scheme distinguishes the two Andrews-Ploberger tests and allows the researcher

Although the two set of tests have been derived with respect to different alternatives, they have limiting distributions that are qualitatively similar. This provides intuition for the fact that these tests have power over a range of alternatives, instead of being limited to the alternatives they are designed for. The limiting distributions of these tests under the random walk alternatives are derived in Stock and Watson (1998) and Elliott and Muller (2003a)[33]. Asymptotic power analysis show that, for local alternatives, all tests perform well and very close to the power envelope. For more distant alternatives, the POI tests perform well. The Nyblom test and the MW test lose power[34], and to a less degree, so do the EW test and the QLR test[35].

4.2.2 The Testing Model and Appropriate Tests

Motivated in Section 3.2, the testing model $y_t = x_t'\beta_t + z_t'\gamma_t + \varepsilon_t$ is concerned with the constancy of β_t, while allowing for other coefficients, γ_t, to vary over time[36]: $H_0 : \beta_t = \beta_0$ and $\gamma_t \neq \gamma_0$ against $H_1 : \beta_t \neq \beta_0$ and $\gamma_t \neq \gamma_0$.

The present testing model is seen to be a non-standard problem because γ_t, the time-varying nuisance sequence, is present under *both* the null *and* the alternative. This distinct feature complicates the testing problem. A desired test, then, should have a limiting representation which is (i) *dependent* on the β_t process and (ii) *independent* of the γ_t process. The dependence property of (i) indicates the test has power against the time variation in β_t. The independence property of (ii) ensures that the instability detected, if any, originates solely from β_t[37]. In what follows, I investigate whether this independence property is possessed by existing stability tests, assuming perhaps erroneously that γ_t is constant.

I first consider the class of tests motivated by discrete breaks. This class includes the QLR test, the MW test and the EW test, which are functionals of sequential Chow statistics. Let SSR_{t_1,t_2} denote the sum of squared residuals from regressing y_t on x_t and z_t over sample period $t_1 \leq t \leq t_2$. The Chow F-statistic testing for a break at date $[sT]$, with $0 \leq s \leq 1$, is

$$F_T(s) = \frac{SSR_{1,T} - SSR_{1,[sT]} - SSR_{[sT+1],T}}{k(SSR_{1,[sT]} + SSR_{[sT+1],T})/(T-k)} \tag{20}$$

to orient the power to certain part of the sample relative to the others. The MW test assigns more weight to alternatives closer to the null and the EW test assigns more weight to alternatives further away from the null.

[33]In Elliott and Muller (2003a), the asymptotic distributions of the POI tests are derived for a general class of breaking processes that cover the random walk process.

[34]This is consistent with the fact that the Nyblom test is designed for local alternatives, and the MW test puts more weight to local alternatives.

[35]See Andrews, Lee and Ploberger (1996), Stock and Watson (1996), Stock and Watson (1998) and Elliott and Muller (2003).

[36]According to the random walk processes for β_t and γ_t specified in Section 3.2, the null hypothesis and the alternative hypothesis can be written equivalently as $H_0 : \lambda_\beta = 0$ and $\lambda_\gamma \neq 0$ against $H_1 : \lambda_\beta \neq 0$ and $\lambda_\gamma \neq 0$.

[37]There are other situations in which a test with such an independence property is desirable. For example, if the goal is to locate the instability among coefficients, then the evidence from a test that is unable to disentangle the instability in the coefficients of interest from that in the nuisance coefficients would not be informative.

where $k = \dim(\beta)$. Let $0 < s_0 < s_1 < 1$, then the QLR, MW and EW statistics are

$$QLR_T = \sup_{s \in (s_0, s_1)} F_T(s); \quad MW_T = \int_{s_0}^{s_1} F_T(s) ds; \quad EW_T = \log \int_{s_0}^{s_1} \exp\left[\frac{1}{2} F_T(s)\right] ds \tag{21}$$

Intuitively this class of tests are independent of γ_t because the Chow F-statistic can be equivalently written as a functional of the estimated difference of β over different subsamples[38]. By Proposition 2, the distribution of $\widehat{\beta}$ is asymptotically independent of γ_t. As a result, the estimator of β in each subsample possesses this independence property, so should their difference.

The second class of tests, designed for random walk alternatives, include the Nyblom test and the POI tests. Consider the Nyblom test first. The standard Nyblom test for the constant γ model $y_t = x_t'\beta_t + z_t'\gamma_0 + \varepsilon_t$ can be constructed in two ways[39],

$$\begin{aligned} L_T^x &= T^{-1} \sum_{t=1}^{T} \left[T^{-1/2} \sum_{i=1}^{t} e_i' x_i' \ \ \widehat{\mathrm{Var}(x_t \varepsilon_t)}^{-1} T^{-1/2} \sum_{i=1}^{t} x_i e_i \right] \\ L_T^u &= T^{-1} \sum_{t=1}^{T} \left[T^{-1/2} \sum_{i=1}^{t} e_i' \widehat{u}_i' \ \ \widehat{\mathrm{Var}(u_t \varepsilon_t)}^{-1} T^{-1/2} \sum_{i=1}^{t} \widehat{u}_i e_i \right] \end{aligned}$$

where e_t is the residual of regressing y_t on x_t and z_t, $\widehat{u}_t$ is the residual of regressing x_t on z_t. L_T^u can be thought of being obtained in two steps: before L_T^u is computed in the second step, the testing model is pre-multiplied by $M_z = I - Z(Z'Z)^{-1}Z'$ to partial out γ_0.

When the constancy of β is tested in model $y_t = x_t'\beta_t + z_t'\gamma_0 + \varepsilon_t$, L_T^x and L_T^u have the same null distribution. They both converge to $\int W(s)'W(s)ds$ where $W(s)$ is a $k \times 1$ standard Brownian motion. But if the true testing model is $y_t = x_t'\beta_t + z_t'\gamma_t + \varepsilon_t$ where the non-tested coefficient γ is unstable, L_T^x and L_T^u behave differently: L_T^u is robust to the instability in γ_t, while L_T^x, unfortunately, is contaminated by γ_t, and hence inappropriate to use for this testing model[40]. The technical insight driven this difference is straightforward: As discussed in Section 3.2,

[38] Chow statistic expressed in the estimated difference of β can be computed as,

$$F_T(s) = \frac{T-k}{k} \left[\widehat{\beta}_{1,[sT]} - \widehat{\beta}_{[sT+1],T}\right]' \left[\frac{1}{s(1-s)} \widehat{\Sigma}_{uu}^{-1} \widehat{\mathrm{Var}(u_t \varepsilon_t)} \widehat{\Sigma}_{uu}^{-1}\right]^{-1} \left[\widehat{\beta}_{1,[sT]} - \widehat{\beta}_{[sT+1],T}\right] \tag{22}$$

where u_t is the residual from the partial regression of x_t on z_t, $\widehat{\Sigma}_{uu}$ and $\widehat{\mathrm{Var}(u_t \varepsilon_t)}$ are the standard estimators for $E(u_t u_t')$ and $\mathrm{Var}(u_t \varepsilon_t)$ in the corresponding stable model. The Chow F-statistic presented in this version is derived under part 1 of Condition 1 that data moments are approximated by the appropriate fractions of the corresponding full-sample moments.

[39] To be more concrete, L_T^x is obtained by computing the Gaussian likelihood ratio of the maximal invariant, $M\epsilon$, under the null and under the alternative, where $M = I - R(R'R)^{-1}R'$ with $R = [X \ Z]$; $\epsilon_t = \varepsilon_t$ under the null of a stable β and $\epsilon_t = \varepsilon_t + x_t'(\beta_t - \beta_0)$ under the alternative of an unstable β. L_T^u can be computed in two steps. In the first step, partial out $z_t'\gamma_0$ by pre-multiplying the matrix form of $y_t = x_t'\beta_t + z_t'\gamma_0 + \varepsilon_t$ by the residual matrix $M_z = I - Z(Z'Z)^{-1}Z'$, which yields $\widetilde{y}_t = \widetilde{u}_t'\beta_t + \widetilde{\varepsilon}_t$, where $\widetilde{y} = M_z y$; $\widetilde{u} = M_z x$ and $\widetilde{\varepsilon} = M_z \varepsilon$. In the second step, L_T^u is obtained by computing the Gaussian likelihood ratio of the maximal invariant, $M_u \widetilde{\epsilon}$, under the null and under the alternative, where $M_u = I - \widetilde{u}(\widetilde{u}'\widetilde{u})^{-1}\widetilde{u}'$, $\widetilde{\epsilon}_t = \widetilde{\varepsilon}_t$ under the null of a stable β and $\widetilde{\epsilon}_t = \widetilde{\varepsilon}_t + \widetilde{u}_t'(\beta_t - \beta_0)$ under the alternative of an unstable β.

[40] To see this, L_T^x is a functional of the partial sum $T^{-1/2} \sum_{i=1}^{t} x_i e_i$. It is shown in the appendix, under

when the stable component of the γ_t process, γ_0, is partialed out in the standard procedure, the time-varying component of the γ_t process, $\gamma_t - \gamma_0$ is also partialed out asymptotically. So the asymptotic effects of γ_t is eliminated *before* L_T^u is constructed in the second step. For this reason, erroneously assuming a constant γ when it is time-varying does no harm. On the other hand, when applying L_T^x to the current testing problem, the tested block $x_t'\beta_t$ and the rest of the model $z_t'\gamma_t + \varepsilon_t$ is not independent because of the correlation between x_t and z_t[41]. Unavoidably, this would lead to an extra term that is non-trivial in the limit, which alters the distribution of the L_T^x statistic.

Similarly, POI tests that are asymptotically independent of the time variation in the non-tested coefficients should be used for testing purpose. Following the same reasoning behind the robust Nyblom test, robust POI tests could be constructed by eliminating the asymptotic effects of γ_t from the regression before computing the test statistics.

In the appendix, it is shown that limiting distributions of QLR_T, MW_T, EW_T and L_T^u are *identical* to their counterparts of the constant γ model, due to the independence of $\widehat{\beta}$ to γ_t. These limiting distributions are first derived in Stock and Watson (1998)[42]. Nevertheless, the asymptotic results are summarized in Proposition 4. Let W and W_β be $k \times 1$ independent standard Brownian motions with $k = \dim(\beta)$. Let $D_B = \mathrm{Var}(u_t\varepsilon_t)^{-1/2}E(u_t u_t')\Sigma_\beta^{1/2}$.

Proposition 4: *Suppose y_t obeys (16) and (17). Suppose Condition 1 holds. Let $h_\lambda(s) = W(s) + \lambda_\beta D_B \int_0^s W_\beta(r)dr$. Then,*

1. $$\widehat{\mathrm{Var}}(u_t\varepsilon_t)^{-1/2}\, T^{-1/2} \sum_{t=1}^{[sT]} \widehat{u}_t e_t \;\Rightarrow\; h_\lambda(s) - s h_\lambda(1)$$
2. $$F_T(s) \;\Rightarrow\; F^*(s) \quad \text{where} \quad F^*(s) = [ks(1-s)]^{-1}[h_\lambda(s) - sh_\lambda(1)]'[h_\lambda(s) - sh_\lambda(1)]$$

As a result of Proposition 4 and the continuous mapping theorem, any test statistic which is a functional of $F_T(s)$, denoted by $\xi_T(F_T(s))$, has the limiting presentation $\xi_T(F^*(s))$[43].

In terms of local asymptotic power, as a result of the independence property of the tests, all findings in the previous literature regarding tests for model $y_t = x_t'\beta_t + z_t'\gamma_0 + \varepsilon_t$, summarized in Section 3.1, carry over to the present testing model $y_t = x_t'\beta_t + z_t'\gamma_t + \varepsilon_t$.

---

both the null hypothesis of a constant β and the alternative of an unstable β, the limiting distribution of $T^{-1/2}\sum_{i=1}^t x_i e_i$ depends on λ_γ and Σ_γ, the two parameters governing the γ_t process. On the other hand, the null and the alternative distributions of $T^{-1/2}\sum_{i=1}^t \widehat{u}_i e_i$ is asymptotically independent of γ_t.

[41] This says, only in the special case that x_t and z_t are asymptotically uncorrelated, L_T^x is robust to the time variation in γ_t. But in general this is not true.

[42] The limiting distributions derived in Stock and Watson (1998) are for testing model $y_t = x_t'\beta_t + \varepsilon_t$, which can be easily extend to the constant-γ model.

[43] In the context of the QLR, MW, EW and the robust Nyblom tests, their asymptotic distribution is given by $L_T^m \Rightarrow \int_0^1 k\ s(1-s)F^*(s)ds$, $QLR_T \Rightarrow \sup_{s\in(s_0,s_1)} F^*(s)$, $MW_T \Rightarrow \int_{s_0}^{s_1} F^*(s)ds$, and $EW_T \Rightarrow \log\int_{s_0}^{s_1} \exp\left[\frac{1}{2}F^*(s)\right] ds$ respectively.

4.3 Size Control of the Joint Test

Since the overall evaluation of the two-step TVP model is a joint test of the stable coefficient restriction (13) and the TVP restriction (14), one issue that has to be addressed is the size of the overall test. If the two individual tests are not independent, then to have a desired significance level of the overall test, the significance levels of both tests have to be adjusted to account for such correlation. Fortunately, this issue does not arise. The following proposition clarifies the statistical relationship between the two tests.

Proposition 5: *Consider the two-step minimum distance problem described by (8)-(10). Suppose Conditions 1 and 2 hold. Then, the standard specification test for restriction $g(\phi_0, \theta_0) = 0$ and the stability tests for restriction $D_g(\phi_t - \phi_0) = 0$ are asymptotically independent under the null hypothesis of $g(\phi_t, \theta_0) = 0$.*

Several remarks regarding Proposition 5 are in order. On the one hand, given the independence, for the joint test to have a correct size, it is easy to fix the size of each of the component tests. On the other hand, the combination of the size of the two tests that matches certain desired size of the overall test is not unique. Different weighting schemes assigned to the two tests would result in different weighted average power for tests of equal (overall) significance level. The standard two-step procedure can be interpreted as the extreme case of assigning all weight to the test for the first restriction $g(\phi_0, \theta_0) = 0$, and hence it has zero power against the alternatives of the second restriction $D_g(\phi_t - \phi_0) = 0$. An equal weight, which indicates the alternatives in the two component tests are equally likely, would result in an overall test with power against alternatives of both component tests[44].

5 Generality of Results to Other Estimation Methods

Most econometric estimators can be viewed as being obtained by minimizing a quadratic form in data and parameters. In the context of estimating a linear Euler equation, the analog between different estimation methods is illustrated in Li (2004). The two-step minimum distance method is, according to Li (2004), (i) essentially the maximum likelihood estimation of the reduced-form model subject to the cross-equation restrictions[45]; and (ii) asymptotically equivalent to

[44]Model selection with parameter instability is also studied in Rossi (2004), according to which, when the alternatives in the two component tests are of equal likelihood, a joint test as the one discussed in the present paper would not achieve the optimal weighted average power. However, an obvious strength of the testing procedure in the current paper is that, if the overall restriction is rejected, the researcher knows exactly which component of the overall restriction fails.

[45]To see this, the OLS estimation of the first-step model is equivalent to the maximum likelihood estimation of the first-step model with Gaussian disturbances. The objective function of the second-step minimum distance

the GMM approach under certain condition[46]. It should not be a surprise that, all results in preceding sections are not limited to linear regressions or two-step minimum distance models. Rather, they extend to other estimation methods.

5.1 Inference on Stable Coefficients

It is seen in Section 3, in the two-step minimum distance model, standard estimation leads to valid inference on the stable coefficients in the presence of reduced-form instability. The same conclusion is obtained in Elliott and Mueller (2003) in linear regression models, and most recently, in Li and Mueller (2004) in the GMM models[47]. Summarizing, "valid inference over stable coefficients" is a general property that applies to a wide range of unstable econometric models. Recall in Section 3.2, the insight of this result was illustrated in a time-varying linear regression model. In the appendix, I show that by treating the linearized GMM first order conditions as generalized "regression" models, this "partialing-out" argument also provides the technical insight for the result of Li and Mueller (2004).

5.2 Model Validation Testing

Section 4 concludes that, in the framework of the two-step model, to evaluate a restriction involving unstable coefficients, one can decompose the restriction into its stable component and its TVP component. A sequential procedure can be applied to test the two components separately. The fact that the test for the stable component is asymptotically independent of the test for the TVP component makes it easy to have an overall test with a desired significance level. In what follows, I extend this statement to other econometric models, by developing an argument applicable to general extremum estimation methods.

Consider the problem of testing a restriction of the generic form

$$a(\phi_t) \;=\; 0 \tag{23}$$

where ϕ_t is a vector of time-varying coefficients, following the TVP process in (9) and (10). $a(\cdot)$ is a possibly non-linear vector function of its argument. The function $a(\cdot)$ is assumed to satisfy some regularity conditions. Let Φ_0 be some neighborhood of ϕ_0.

estimation is the maximum likelihood Wald statistic of testing the cross-equation restrictions.

[46] As shown in Li (2004), when the first-step model is a reduced-form vector autoregression, the two-step method and GMM are asymptotically equivalent as long as the GMM instrument set is identical to the VAR information set.

[47] Although different from the analysis in the present paper, Li and Mueller (2004) assume the time-varying coefficients follow a deterministic path, rather than a random process. Many other papers in the literature adopt this assumption to study parameter instability. See for example, Ghysels and Hall (1990a), Andrews and Ploberger (1994), Sowell (1996), Rossi (2003), among others.

Condition 3: *$a(\phi)$ and $\partial a(\phi)/\partial\phi$ are continuously differentiable in ϕ for all $\phi \in \Phi_0$ with* $\sup_{\phi\in\Phi_0} \|\partial^2 a^{(i)}(\phi)/\partial\phi\partial\phi'\| < \infty$ *for all $i \in [1, ..., dim(a)]$.*

Let $Q_T(\phi)$ be the objective function in estimating ϕ_0. It could be the objective function of OLS, NLS, MLE or GMM, among others. Denote the first and the second derivatives of $Q_T(\phi)$ with respect to ϕ as $\partial Q_T(\phi)/\partial\phi = T^{-1}\sum s_t(\phi)$ and $\partial^2 Q_T(\phi)/\partial\phi\partial\phi' = T^{-1}\sum h_t(\phi)$, where $s_t(\phi)$ and $h_t(\phi)$ are the score and Hessian functions for date t observation[48]. To establish the asymptotic results, the score and Hessian functions are assumed to satisfy Condition 4 below. Note that Condition 4 is a generalized version of Condition 1[49], so that it not only applies to linear models, but also covers a wide variety of non-linear models. In Condition 4 below, let W_ϕ be a standard Brownian motion associated with ϕ_t.

Condition 4: *Functions $s_t(\phi)$ and $h_t(\phi)$ satisfy the following requirements under ϕ_t.*

1. $T^{-1}\sum_{t=1}^{[sT]} h_t(\phi) \xrightarrow{p} sEh_t(\phi)$ *uniformly for all $\phi \in \Phi_0$.*

2. $T^{-1/2}\sum_{t=1}^{[sT]} s_t(\phi_t) \Rightarrow F(s)$ *for some mean zero Gaussian stochastic process $F(s)$ that has $E[F(s)F(s)'] = s\ \Omega$ where Ω is a positive definite matrix.*

3. $T^{-1}\sum_{t=1}^{[sT]} h_t(\phi_0)T^{-1/2}\sum_{i=1}^{t}\nu_i \Rightarrow Eh_t(\phi_0)\Sigma^{1/2}\int_0^s W_\phi(r)dr$

Let $\widehat{\phi}$ and $\widetilde{\phi}$ denote the unconstrained and constrained extreumum estimators of ϕ_0 obtained by ignoring the instability in ϕ_t, so that $\widehat{\phi} = \text{argmin}_{\phi\in\Phi_0} Q_T(\phi)$ and $\widetilde{\phi} = \text{argmin}_{\phi\in\Phi_0} Q_T(\phi)$ subject to $a(\phi) = 0$. Let $V_{\widehat{\phi}}$ and $V_{\widetilde{\phi}}$ be the asymptotic variances of $\widehat{\phi}$ and $\widetilde{\phi}$ in the corresponding stable model. Let $LR_T = 2T[Q_T(\widetilde{\phi}) - Q_T(\widehat{\phi})]$, $LM_T = T\gamma_T'\widehat{\text{Avar}}(\gamma_T)^{-1}\gamma_T$, and $W_T = Ta(\widehat{\phi})'\widehat{\text{Avar}}(a(\widehat{\phi}))^{-1}a(\widehat{\phi})$ be the likelihood ratio test[50], Lagrange multiplier test and Wald test statistics, respectively, and γ_T is the Lagrange multiplier in the constrained problem, $\widehat{\text{Avar}}(a(\widehat{\phi}))$ and $\widehat{\text{Avar}}(\gamma_T)$ are the estimated asymptotic variances of $a(\widehat{\phi})$ and γ_T assuming a stable model[51]. Parameters λ and Σ are defined in (9) and (10). Then, I have the following proposition that parallels the set of testing results for the two-step minimum distance model.

[48] For estimation methods other than MLE, calling $s_t(\phi)$ and $h_t(\phi)$ the score and Hessian functions is a slight abuse of the language because $Q_T(\phi)$'s for methods other than MLE are not the likelihood function. But I will continue to use these terms because of the analog between various methods in deriving the asymptotic results.

[49] Consider regression model $y_t = x_t'\phi_t + \varepsilon_t$, then $Q_T(\phi) = T^{-1}\sum e_t(\phi)'e_t(\phi)$ where $e_t(\phi) = y_t - x_t'\phi$, $s_t(\phi) = x_t\varepsilon_t$, and $h_t(\phi) = x_t x_t'$. When $h_t(\phi)$ and $s_t(\phi)$ in Condition 4 is replaced by their corresponding OLS expressions, we obtain Condition 1. Moreover, similar to Condition 1, any primitive assumptions that lead to Condition 4 *under* ϕ_0, together with the *contiguity* argument would be sufficient to guarantee Condition 4, which is defined *under* ϕ_t.

[50] For the statistic LR_T, it might be more appropriate to call it *distance metric* test, rather than likelihood ratio test, for methods other than maximum likelihood estimation. But for simplicity, the name "likelihood ratio test" is used for all estimation methods.

[51] By the theory results in Section 2.3, as long as (i) $\widehat{\text{Avar}}(a(\widehat{\phi}))$ and $\widehat{\text{Avar}}(\gamma_T)$ are consistent estimator in the

Proposition 6: *Consider the testing problem described by (9), (10) and (23). Under Conditions 3 and 4, following results hold.*

1. *$\widehat{\phi}$ and $\widetilde{\phi}$ are consistent and asymptotically normal. Their limiting distributions are given by $T^{-1/2}(\widehat{\phi} - \phi_0) \Rightarrow \mathcal{N}(0,\ V_{\widehat{\phi}} + \frac{1}{3}\lambda^2\Sigma)$ and $T^{-1/2}(\widetilde{\phi} - \phi_0) \Rightarrow \mathcal{N}(0,\ V_{\widetilde{\phi}} + \frac{1}{3}\lambda^2\Sigma)$.*
2. *Evaluation of $a(\phi_t) = 0$ can be proceeded by sequentially testing the TVP restriction $A(\phi_0)(\phi_t - \phi_0) = 0$, where $A(\phi_0) = \partial a(\phi_0)/\partial\phi$, and the stable restriction $a(\phi_0) = 0$.*
3. *Standard specification tests for $a(\phi_0) = 0$, including the standard likelihood ratio test, the Lagrange multiplier test and the Wald test, are independent of the time variation in ϕ_t, under the null of $a(\phi_t) = 0$. Test statistics LR_T, LM_T and W_T all converge to a central χ^2 distribution, with degrees of freedom equal to the dimension of $a(\phi_t)$.*
4. *Standard specification tests for $a(\phi_0) = 0$ are asymptotically independent of Chow-statistic based stability tests for $A(\phi_0)(\phi_t - \phi_0) = 0$ under the null of $a(\phi_t) = 0$.*

According to part 1 of Proposition 6, the conventional estimators of ϕ_0 remain consistent in the presence of instability, but standard deviations of elements of $\widehat{\phi}$ or $\widetilde{\phi}$ based on the conventional distributions will be incorrect. The standard t-test or F-test for the coefficients will lead to over-rejection in general. Also, note that, the distortionary effect of ϕ_t solely depends on λ and Σ, the two parameters governing the ϕ_t process, while other model features are irrelevant to the magnitude of the distortion[52]. The effects of parameter instability on inference are also studied in Li and Mueller (2004) for GMM models. Due to the deterministic nature of the time-varying coefficients assumed in Li and Mueller (2004), instability induces a constant shift, rather than a change in the variance, in the limiting distributions of the standard estimators, and the size of the bias only depends on the path of ϕ_t. Therefore, part 1 of Proposition 6 is complementary to that of Li and Mueller (2004).

According to part 3 of Proposition 6, although distribution of $\widehat{\phi}$ or $\widetilde{\phi}$ is distorted by the ignored instability in ϕ_t, distributions of the standard specification tests remain unaltered. A restatement of this result is, instability would easily remain undetected with the standard tests because the standard tests have no power in distinguishing a class of local time-varying alternatives of the kind considered in the paper. In the literature, a similar point has been made in Newey (1985), Ghysels and Hall (1990a), and most recently, Li and Mueller (2004), which show that in GMM models, Hansen's J-test has no asymptotic power against a class of deterministic time variations in the coefficients over the sample period. Though the above mentioned studies, including the current paper, consider different forms of local alternatives against the null of stability, essentially the problems reported for the standard specification

corresponding stable model, and (ii) contiguity holds for the data generating process, these estimators remains consistent in the unstable model.

[52] So the distortion on the variance of $\widehat{\phi}$ reported in (15) for the linear regression model is just a special case of the first result in Proposition 6.

tests can all be classified as *identification* problems, because in all these studies, the extra quadratic term induced by the time varying coefficients which would lead to a non-central χ^2 distribution of the standard tests, turns out to be zero for certain class of local instabilities due to various kinds of singularity[53].

It follows that, as stated in part 2 of proposition 6, standard specification tests alone are not adequate for testing an overall restriction with potentially unstable coefficients. Instead, they must be supplemented by a stability assessment of the restriction. Actually, this issue has been raised in the literature long before the present paper. See, for example, Ghysels and Hall (1990a and 1990b) and Oliner, Rudebusch and Sichel (1996). Largely motivated by the rationale of the Lucas (1976) Critique, in their discussions about GMM model validation tests of Euler equations, the authors stress that Structural stability tests are a natural diagnostic for Euler conditions. In the present paper, this intuition is formalized econometrically and it is seen to apply to a broad range of econometric models. The analysis of the paper also enables me to clarifies the statistical relation between the standard specification tests and the stability tests. By part 4 of Proposition 6, stability tests should not be viewed as substitutes for standard specification tests, and vice versa. Rather, they are complementary to, and independent of, one another.

In practice, the necessity of stability tests as a model diagnostic appears little received. Most empirical work in macroeconomics literature judges model adequacy of Euler conditions using *only* the conventional specification tests. According to the results of the paper, with a changing economic environment, if an Euler equation fails the standard specification test, it fails the overall model assessment. But, if an Euler equation is not rejected by the standard specification test, it does *not* necessarily imply it is a valid structural representation. In this sense, to determine whether they are truly structural, those Euler equation models having been received supporting evidence from the standard specification tests should be submitted to further scrutiny by testing whether they exhibit structural stability. Such an example is provided in the application part of the paper. One Euler equation widely accepted in empirical literature on the basis of the standard specification test, fails the stability assessment. Though any empirical exercise that ignores the potential time variation in the data generating process can be interpreted as subjectively assuming stability holds with probability one, given the widespread instability found in macroeconomic relations, this seems an inappropriate weight

[53] In the current paper, the family of instabilities that (i) take the TVP form in (9) with the nesting in (10), and (ii) moves along directions characterized by $\frac{\partial a(\phi_0)}{\partial \phi}\phi_t = \frac{\partial a(\phi_0)}{\partial \phi}\phi_0$, will lead to a zero non-centrality parameter. Here, singularity can be seen to arise from $\mathrm{Var}(\phi_t)$, or equivalently, according to (9) and (10), arise from the matrix Σ, which has a reduced rank of $dim(\phi) - dim(a(\phi))$. In Newey (1985), the family of deterministic local alternatives that would cause the consistency problem in GMM testing is characterized by setting the non-centrality parameter of the distribution of the J-test to zero. In Ghysels and Hall (1990), the instability in the form of a single discrete break turns out to be a particular member of the class of alternatives characterized in Newey (1985).

scheme assigned to the two parts of the overall model validation. Therefore, the analysis of the paper recommend a major re-orientation in the evaluation of macroeconomic models: stability tests should routinely be reported, in addition to the conventional specification tests.

6 An Application to Investment Euler Equations

6.1 Background

Since Lucas' (1976) indictment on several traditional models including the traditional investment model, the subsequent research program of rational expectations has made Euler equation a popular approach to modeling investment. Best known early theoretical work on firm's optimal investment decision includes Hayashi (1982) and Abel and Blanchard (1983), which derive the investment Euler from firm's profit maximization subject to technology and adjustment cost of investment. Since then this formulation of investment decision has become standard and showed up in mainstream macroeconomics textbooks, for instance, Blanchard and Fischer (1989), Romer (1996) and Obstfeld and Rogoff (1996). In recent literature, most theoretical work that involves endogenous investment focus on general equilibrium analysis, with firm's investment decision being part of a fully-structural model. But the modeling of the investment block remains the standard Euler equation approach, examples are King and Watson (1996), Jermann (1998), Bernanke, Gertler and Gilchrist (1998), Kim (1999), Casares and McCallum (2000), Edge (2000), Christiano, Eichembaum and Evans (2001), Edge, Laubach, and Williams (2003) and Smets and Wouters (2003a, 2003b), among many others.

There is a relatively smaller empirical literature on investment Euler equations. Abel (1980) was probably the first paper to estimate an investment Euler equation with rational expectations. Later work includes Pindyck and Rotermberg (1983), Shapiro (1986a, 1986b), Gilchirist (1990), Gertler, Hubbard and Kashyap (1991), Hubbard and Kashyap (1992), Whited (1992), Oliner, Rudebusch and Sichel (1995 and 1996) and Gilchrist and Himmelberg (1998). Generalized method of moments (GMM) is used in all of these papers to estimate the structural coefficients.

In terms of model validation testing, empirical work cited above mostly obtained supporting results. The supporting empirical evidence was found on the basis of Hansen's J-test for over-identifying restrictions *only*[54]. Unfortunately, as discussed in preceding sections, standard specification tests are not enough for overall model evaluation. I have shown in theory that when

[54] In general equilibrium empirical models such as those estimated in Christiano, Eichembaum and Evans (2001) and Smets and Wouters (2003a and 2003b), typically the empirical adequacy of the investment block is not investigated separately. Instead, it is only a part of the evaluation of the general equilibrium model. Again, there is little examination of stability of the estimated parameters of taste and technology governing objective functions.

there is instability in the data generating process, rather than relying on conventional specification tests as the sole model diagnostic, stability assessment of structural relations should be an essential part of an overall model evaluation. This is consistent with the standards of the Lucas (1976) Critique, according to which, such local alternatives are precisely the relevant ones to confront when testing the success of estimated Euler equations. Intuitively, if economic theory predicts a structural equation to be invariant across time and regimes, then the estimated equation should exhibit a similar invariance, suppose the structural model is correctly specified. In this aspect, to my knowledge, Oliner, Rudebusch and Sichel (1996) is probably the only paper to date that incorporates stability study into the overall evaluation of an investment Euler equation, and their results point to parameter instability, and hence inadequacy of the investment model they studied.

In this section, using the two-step classical minimum distance method, I examine several investment models that are typical of those used in the macroeconomic literature. The motivation of this examination is twofold. First, I aim to conduct a relatively more comprehensive analysis in investment models than those having been reported in the literature. To be more concrete, three investment Euler equations are considered, with (1) investment-capital ratio, (2) capital growth, and (3) investment growth being the decision variable respectively. Among them, to my knowledge, evaluation of the model in investment growth and the model in capital growth has not been reported in the empirical literature. The model in investment-capital ratio, on the other hand, is arguably the most popular model in both theoretical and empirical macroeconomics. For this model, my aim is to determine whether evidence from my study, especially from various stability tests contradicts that from standard specification tests in the previous literature. In this paper, these three investment models are diagnosed using the conventional J-test for the cross-equation restrictions the investment Eulers imposed on the reduced-form models, together with a battery of structural stability tests.

A second motivation is, as documented in Li (2004), the two-step minimum distance estimator is essentially a ML estimator of the reduced-form model subject to the cross-equation restrictions, while almost all previous empirical research on investment Euler equations was conducted in the GMM framework. ML estimator and GMM estimator behave very differently in finite samples[55]. Simulation study in Fuhrer, Moore and Schuh (1995) suggests that the small sample performance of ML estimator is superior to that of GMM estimator in various aspects[56]. Thus, in the current context of investment Euler equations, the two-step minimum distance estimator is expected to possess better finite sample properties than the most widely-used GMM

[55] It is well known in the GMM framework, pathological cases caused by weak instruments are common and shown to impart serious distortion on inference and specification tests. See the series of papers on the finite sample performance of GMM estimators in *Journal of Business and Economic Statistics*, 14, 1996. For example, Altonji and Segal (1996) and Hansen, Heaton and Yaron (1996).

[56] In Fuhrer, Moore and Schuh (1995), small sample properties of a ML estimator and a GMM estimator are compared in the context of a linear-quadratic inventory Euler equation.

estimator, and hence, to provide more reliable empirical evidence.

6.2 Investment Equations

The investment models examined in the paper are some standard investment Euler equations derived from the representative firm's profit optimization. Although the assumptions underlying the model are open to criticism, I have adopted them because I intend to assess the validity of *standard* models that have been used in the literature.

6.2.1 Common features

Investment Euler equations based on adjustment cost functions have become a fixture in applied work. The investment models considered in the paper differ only in their specifications of the adjustment cost of capital stock. In what follows, I first present the common features shared by these models before introducing various specifications on the adjustment cost. The firm's production function is assumed to be of Cobb-Douglas with constant returns to scale,

$$Y_t = F(K_t, L_t) = A_t K_t^{\alpha} L_t^{1-\alpha}, \tag{24}$$

where K_t is the capital stock at the beginning of period t; L_t is the employment during period t; A_t is an exogenous shifter of the total productivity; and the capital share in production, α, is a constant with $0 < \alpha < 1$. Capital is subject to adjustment cost. Let C_t be the adjustment cost of capital in period t. Following the usual practice in the literature, the representative firm is assumed to maximize the expected present value of real future profits

$$E_t\left[\sum_{s=t}^{\infty} \beta^{s-t} R_s\right]. \tag{25}$$

where β is the firm's time discount factor which is assumed to be constant. The period real profit of the firm, R_s, is defined as[57]

$$R_s = F(K_s, L_s) - C_s - \omega_s L_s - I_s p_s^I \tag{26}$$

where I_s is the gross investment during period s, p_s^I is the relative purchase price of capital goods, and ω_s is the real wage in period s. The capital stock, K_t, evolves according to the following law of motion,

$$K_{t+1} = (1-\delta)K_t + I_t, \tag{27}$$

where δ is a constant depreciation rate on the firm's capital stock.

[57] Rather than introducing the adjustment costs in the period profit function, an alternative way to introduce the adjustment costs is to place it in the law of motion for capital in (27). The two formulations are shown to deliver similar results concerning the optimal investment rule, see Hayashi (1982). I henceforth focus on the current formulation.

6.2.2 Adjustment costs of capital

At the aggregate level, changes in capital stock occur at a finite rate and investment is persistent. The persistence and the smooth and continuous changes in investment prompted researchers to focus on *convex* adjustment costs. In applied work, adjustment costs are typically assumed to be quadratic. Specifically, two classes of specifications used in the literature are considered. The first class of cost functions model the adjustment cost as a function of *investment-capital* ratio. This class of cost functions are, thus far, most widely used in the literature[58],

$$C(I_t, K_t) = [\phi_1(I_t/K_t) + \phi_2(I_t/K_t)^2]K_t \tag{28}$$

$$C(I_t, K_t) = [\phi_1(I_t/K_t) + \phi_2(I_t/K_t)^2]I_t \; p_t^I. \tag{29}$$

Both specifications penalize the investment share in capital stock I_t/K_t[59]. For (28) and (29) to be well-defined, ϕ_2 in both equations should be positive[60].

Another class of cost functions that have also been used in the literature are as follows[61],

$$C(I_t, I_{t-1}) = \left[\phi_0 + \phi_1(I_t/I_{t-1}) + \phi_2(I_t/I_{t-1})^2\right] I_t \; p_t^I \tag{30}$$

$$C(I_t, I_{t-1}) = \widetilde{\phi}_1 I_t + \widetilde{\phi}_2 I_t^2 + \widetilde{\phi}_3(I_t - I_{t-1}) + \widetilde{\phi}_4(I_t - I_{t-1})^2. \tag{31}$$

Specification (30) penalizes the growth in investment. Specification (31) penalizes *both* the level of investment *and* the change in investment. For cost functions (30) and (31) to be well-defined, ϕ_2, $\widetilde{\phi}_2$ and $\widetilde{\phi}_4$ should all take positive values[62].

[58] See, for example, Hayashi (1982), Abel and Blanchard (1983), Shapiro (1986a and 1986b), Blanchard and Fischer (1989), Gilchrist (1990), Hubbard and Kashyap (1992), Whited (1992), Baxter and Crucini (1993), Oliner, Rudebusch and Sichel (1995 and 1996), King and Watson (1996), Romer (1996), Obstfeld and Rogoff (1996), Bernanke, Gertler and Gilchrist (1998), Jermann (1998), Kim (1999), Edge (2000) and Casares and McCallum (2000), among many others.

[59] The only difference is that cost function (28) is expressed in the unit of capital goods while cost function (29) is expressed in the unit of investment.

[60] This is because marginal adjustment costs are supposed to be increasing with the investment-capital share, which implies $\phi_2 > 0$.

[61] See for example, Christiano, Eichembaum and Evans (2001), Smets and Wouters (2003a, 2003b) and Edge, Laubachard and Williams (2003). It should be pointed out, (i) the specification used in Edge, Laubachard and Williams (2003) is $C(I_t, I_{t-1}) = \exp\{\phi_1[(I_t/I_{t-1} - \phi_2)]^2\}I_t p_t^I$. Since my empirical exercise in this section is based on the log-linearized equations, this specification results in an identical log-linearized investment Euler as cost function (30); (ii) According to the analysis in Christiano, Eichembaum and Evans (2001), in a general equilibrium setup, a cost function penalizing the change in investment, such as specification (30) and specification (31), generates theoretical impulse responses to a monetary policy shock that better fit the empirical finding in investment than other cost specifications.

[62] In (30), the marginal adjustment cost is assumed to be increasing with the change in investment, which implies $\phi_2 > 0$; in (31), the marginal adjustment costs are supposed to be increasing with *both* the level *and* the change of investment, which implies $\widetilde{\phi}_2 > 0$ and $\widetilde{\phi}_4 > 0$.

6.2.3 Investment Euler equations

The firm's optimization problem is then to choose processes I_t, K_t and L_t for all dates $t \geq 0$ to maximize (25) subject to (24), (26), (27) and one of the cost functions from (28) to (31). This is a standard optimization problem subject to a sequence of constraints. Different specifications of the adjustment cost in capital lead to different investment Euler equations. The log-linearized first-order condition based on cost function (28) or (29) takes the form

$$\widehat{IK}_t = \beta E_t \widehat{IK}_{t+1} + \omega_1[\beta(1-\delta)E_t\widehat{p}^I_{t+1} - \widehat{p}^I_t] + \omega_2 E_t \widehat{KY}_{t+1} \tag{32}$$

where a caret over a variable signifies the variable's log-deviation from its steady-state value. $\widehat{IK}_t = \widehat{I}_t - \widehat{K}_t$ is the logarithmic investment share in capital stock and $\widehat{KY}_t = \widehat{K}_t - \widehat{Y}_t$ is the logarithmic capital share in total output. Using the log-linearized version of the law of motion of capital in (27), equation (32) can be rewritten in terms of capital growth[63],

$$\Delta\widehat{K}_t = \beta E_t \Delta\widehat{K}_{t+1} + \widetilde{\omega}_1[\beta(1-\delta)E_t\widehat{p}^I_{t+1} - \widehat{p}^I_t] + \widetilde{\omega}_2 E_t \widehat{KY}_{t+1} \tag{33}$$

where $\Delta\widehat{K}_t = \widehat{K}_{t+1} - \widehat{K}_t$ is the growth rate of capital in period t[64]. The investment Euler equation based on cost function (30) is

$$\Delta\widehat{I}_t = \beta(2-\delta)E_t\Delta\widehat{I}_{t+1} - \beta^2(1-\delta)E_t\Delta\widehat{I}_{t+2} + \gamma_1\widehat{p}^I_t + \gamma_2 E_t\widehat{p}^I_{t+1} + \gamma_3 E_t\widehat{p}^I_{t+2} + \gamma_4 E_t\widehat{KY}_{t+1} \tag{34}$$

where $\Delta\widehat{I}_t = \widehat{I}_t - \widehat{I}_{t-1}$ is the percentage change in investment. Finally, cost function (31) results in the following first order condition

$$\begin{aligned} \Delta\widehat{I}_t \quad = \quad & \beta(2-\delta)E_t\Delta\widehat{I}_{t+1} - \beta^2(1-\delta)E_t\Delta\widehat{I}_{t+2} \\ & +\lambda_1[\beta(1-\delta)E_t\widehat{I}_{t+1} - \widehat{I}_t] + \lambda_2[\beta(1-\delta)E_t\widehat{p}^I_{t+1} - \widehat{p}^I_t] + \lambda_3 E_t\widehat{KY}_{t+1} \end{aligned} \tag{35}$$

Derivation of the Euler equations is included in the appendix. Since the coefficients ω's, $\widetilde{\omega}$'s, γ's and λ's are functions of (i) the so-called "deep" structural parameters of taste and technology in equations (24) to (27), and (ii) steady-state values of the variables, they are supposed to be stable across time and regimes, if the objective function and the structural constraints are correctly specified.

Note that in equations (32), (33) and (35), $\beta(1-\delta)$ is very close to one for theoretical values of β and δ, which justifies the use of $E_t\Delta\widehat{I}_{t+1}$ and $E_t\Delta\widehat{p}^I_{t+1}$ to approximate $\beta(1-\delta)E_t\widehat{I}_{t+1} - \widehat{I}_t$ and $\beta(1-\delta)E_t\widehat{p}^I_{t+1} - \widehat{p}^I_t$ respectively[65]. Similarly, in equation (34), by expressions of the γ's

[63] Although (32) and (33) are mutually equivalent models in economic theory, empirical results could be very different for the two specifications, say, due to a misspecified law of motion of capital and inaccurate data.

[64] As defined in equation (24), K_{t+1}, hence the log-linearized $\widehat{K}_{t+1}$, is a date t variable since it denotes the capital stock accumulated by the end of period t. Here, in equation (33), to avoid confusion, I use $\Delta\widehat{K}_t$ instead of $\Delta\widehat{K}_{t+1}$ to denote the capital growth by the end of period of t.

[65] For $\beta = 0.99$ and $\delta = 0.025$, $\beta(1-\delta) = 0.97$, $\beta(1-\delta)E_t\widehat{I}_{t+1} - \widehat{I}_t = E_t(0.97\widehat{I}_{t+1} - \widehat{I}_t)$. Note that $0.97\widehat{I}_{t+1} - \widehat{I}_t$ is too highly correlated with $\widehat{I}_{t+1} - \widehat{I}_t$. This may induce severe statistical imprecision of the estimation results. Same argument applies to $\beta(1-\delta)E_t\widehat{p}^I_{t+1} - \widehat{p}^I_t$.

provided in the appendix, $\gamma_1+\gamma_2+\gamma_3$ is close to zero for theoretical values of β and δ, which implies $\gamma_1\widehat{p}^I_t+\gamma_2E_t\widehat{p}^I_{t+1}+\gamma_3E_t\widehat{p}^I_{t+2}$ can be approximated by $-\gamma_1E_t\Delta\widehat{p}^I_{t+1}+\gamma_3E_t\Delta\widehat{p}^I_{t+2}$. With these approximations, in the rest of the section, I estimate and test the following three specifications,

$$\begin{aligned}
\widehat{IK}_t &= \theta_{IK}E_t\widehat{IK}_{t+1}+\theta_{p1}E_t\Delta\widehat{p}^I_{t+1}+\theta_{KY}E_t\widehat{KY}_{t+1}+\varepsilon_t \\
\Delta\widehat{K}_t &= \theta_{K}E_t\Delta\widehat{K}_{t+1}+\theta_{p1}E_t\Delta\widehat{p}^I_{t+1}+\theta_{KY}E_t\widehat{KY}_{t+1}+\varepsilon_t \\
\Delta\widehat{I}_t &= \theta_{I1}E_t\Delta\widehat{I}_{t+1}+\theta_{I2}E_t\Delta\widehat{I}_{t+2}+\theta_{p1}E_t\Delta\widehat{p}^I_{t+1}+\theta_{p2}E_t\Delta\widehat{p}^I_{t+2}+\theta_{KY}E_t\widehat{KY}_{t+1}+\varepsilon_t.
\end{aligned} \tag{36}$$

where the first equation in (36) is a modified, or an approximate, version of (32), the second equation is a modified version of (33), and the third equation nests a modified version of (34) and a modified version of (35). Latter in the paper, the three specifications in (36) are referred to as "the ΔI model", "the IK model" and "the ΔK model" respectively. A disturbance term ε_t, which does not belong to original Euler equations, is added to each of the equations. Different interpretation of ε_t is possible. It may capture misspecification of the models or approximation errors induced by linearizion.

A common feature of the investment equations in (36) is that all specifications are purely forward-looking equations that relate a decision variable, which could be the investment-capital ratio ($\widehat{IK}$), the capital growth ($\Delta\widehat{K}$) or the investment growth ($\Delta\widehat{I}$), to expected future values of that decision variable, changes in the purchase price of capital goods ($\Delta\widehat{p}^I$) and capital share in output ($\widehat{KY}$). Moreover, economic theory predicts the signs, and in some cases the values, of coefficients of these investment models, as summarized in Table 8 in the appendix.

6.3 Econometric Specifications

The first-step model I use to summarize the dynamics of the economy is a pth-order reduced-form vector autoregression. Let y_t denote an n-variable vector containing variables in an investment equation and other useful variables summarizing the behavior of the economy. Let $Z_t=[y_t \;\; y_{t-1} \;\; \cdots \;\; y_{t-p+1}]$. Z_t is the econometrician's information set at date t. If the reduced-form model were time-invariant, the vector autoregression in the first-order form would be $Z_t=\Phi_0Z_{t-1}+u_t$ where the error term u_t $E(u_t|Z_{t-1})=0$.

Then the assumption of $E(\varepsilon_t|Z_{t-1})=0$, that is, the disturbance in an investment equation at date t is not predictable at date $t-1$, leads to a vector of standard cross-equation restrictions, in which all coefficients are stable. For example, the cross-equation restriction for the model in ΔI in a stable environment is

$$0 \;=\; \left[I-\Phi_0'\theta_{I1}-(\Phi_0')^2\theta_{I2}\right]e_I-\left[\Phi_0'\theta_{p1}-(\Phi_0')^2\theta_{p2}\right]e_p-\Phi_0'\theta_{KY}e_{KY}=0,$$

where e_I, e_p and e_{KY} are corresponding $np\times 1$ selection vectors for $\Delta\widehat{I}$, $\Delta\widehat{p}^I$ and $\widehat{KY}$ in the vector autoregression[66]. See the appendix for the derivation.

[66] To be more specific, in a first-step VAR, suppose $\Delta\widehat{I}$, $\Delta\widehat{p}$ and $\widehat{KY}$ take the first, the second and the third positions in Z_t respectively, then $e_I=[1\;0\;0\cdots0]'$, $e_p=[0\;1\;0\cdots0]'$, $e_{KY}=[0\;0\;1\cdots0]'$. For the model in IK and the model in ΔK, selection vectors e_{IK} and e_K in (37) are defined in the same way.

However, any single source of instability in the economic environment, say, changing market conditions and economic policies, can translate itself into the instability in the reduced-form coefficients. Indeed, the instability of the VARs used in this application is confirmed later in this section. Thus, the appropriate first-step model is an *unstable* vector autoregression

$$Z_t = \Phi_t Z_{t-1} + u_t$$

where Φ_t is assumed to follow the TVP process specified in (9) and (10). With the first-step coefficient matrix Φ_0 changing to Φ_t, the cross-equation restrictions change accordingly. The resulting time-varying coefficients cross-equation restrictions for the three models in (36) are

$$\begin{aligned}
0 &= [I - \Phi_t'\theta_{IK}]\, e_{IK} - \Phi_t'\theta_{p1} e_p - \Phi_t'\theta_{KY}\ e_{KY} \qquad (37)\\
0 &= [I - \Phi_t'\theta_K]\, e_K - \Phi_t'\theta_{p1} e_p - \Phi_t'\theta_{KY}\ e_{KY}\\
0 &= \left[I - \Phi_t'\theta_{I1} - (\Phi_t')^2\theta_{I2}\right] e_I - \left[\Phi_t'\theta_{p1} - (\Phi_t')^2\theta_{p2}\right] e_p - \Phi_t'\theta_{KY}\ e_{KY}
\end{aligned}$$

respectively. Estimation and model validation can then be conducted applying the procedure developed in previous sections. Estimation and testing procedure in the context of the three investment equations in (37) is summarized in the appendix.

6.4 Empirical Results

6.4.1 Data, persistence and reduced-form instability

The estimates of the this section use quarterly data for the United States from 1967:I to 2001:IV. In addition to the variables in an investment equation, I also include other variables, which may help predict the decision variable of that investment equation, into the first-step vector autoregression model. For instance, output is included in all VAR specifications to capture the effects of output movement. This could be justified by the so-called "acceleraionist" theory of investment that relates investment to changes in output. The data appendix documents the construction and data source for each series. Two schemes are considered to formulate the vector autoregression. In the first scheme, five variables are included, which are (1) the decision variable of an investment Euler equation, (2) purchase price of capital goods, (3) capital share, (4) a short-term interest rate and (5) total output. In the second scheme, all variables in the first scheme are included *except for* the short-term interest rate. This allows me to study the sensitivity of my estimation results to the inclusion/exclusion of the interest rate. To further check the robustness of the empirical results, within each scheme, different measures of variables and different numbers of lags are considered. See the data appendix for documentation.

One feature of the data that deserves attention is the persistence of the time series. In the context of the two-step minimum distance procedure, even if the first-step model is time-invariant, highly persistent series entering the first-step VAR could lead to inference problem in the second-step[67]. In the presence of instability, the compound effects of time-varying coefficients and persistent regressors could only deteriorate the problem. Persistence of the series and corresponding transformations made to the series to circumvent the problem are documented in the data appendix.

Before proceeding to the results on structural investment equations, stability of the first-step VARs must be addressed to justify the TVP specification and the procedure developed for it in the current application. As documented in the appendix, persistent but small instability is found in *all* VAR specifications. Thus empirical evidence suggests the TVP model (9) and the asymptotic nesting (10) are appropriate[68].

6.4.2 Estimation and testing results for investment models

Next let's turn to the estimation and validation of the second-step model, the structural investment equations. Three criteria are used to evaluate a model. An investment model is regarded as not being rejected by the data if (i) the conventional J-test does not reject the cross-equation restrictions; (ii) the stability requirement of the cross-equation restrictions is satisfied; and (iii) the estimated coefficients have signs consistent with economic theory. The first two criteria judge an estimated equation quantitatively by the standards of statistical tests. The third criterion evaluates estimated coefficients in a qualitatively manner, guided by economic theory.

Table 3 to Table 5 present the estimates for models in ΔI, in IK and in ΔK respectively, accompanied by the J-test and stability tests. Table 2 summarizes the results in Table 3 to Table 5 in a qualitative manner. Inspection of Table 2, *none* of the estimated Euler equations satisfy all three criteria simultaneously. The estimated IK model and the estimated ΔK model yield expected signs on all coefficients. However, they fail one or both of the J-test and stability tests. In particular, the IK model, for which previous research have obtained supporting evidence from GMM over-identifying tests, is seen to exhibit instability, and this observed instability turns out to be a result robust to various VAR specifications. Moreover, compared to the corresponding IK equation in the reduced-form VAR, the Euler equation appears to provide no more improvement when judged on the basis of structural stability.

[67] It is well known that time series regression that includes persistent variables can behave very differently than a standard regression. In the two-step minimum distance model, the effects of persistence extends to the second-step because the second-step estimator can be written as linear combinations of the first-step VAR coefficients. See Li (2004) for a more detailed analysis on this issue. In the context of an Euler equation for inflation, Li (2004) shows that highly persistent variables can lead to false rejections.

[68] In addition to persistence and instability, the other data issue is that VAR errors might exhibit conditional heteroskedasticity and autocorrelation. Temporary time variations in the VAR coefficients could also be thought of as part of the heteroskedastic and autocorrelated error, as discussed in Section 2.2. Failing to account for heteroskedasticity and autocorrelation when it is present affects the optimal weighting matrix in the second step, contaminating standard deviations and test statistics. This motivates using the robust version of estimation since it imposes less restriction on the data. In this section, only robust results are reported.

Table 2: Evaluation of Investment Euler Equations

	Model in ΔI		Model in IK		Model in ΔK	
	VAR with i	VAR without i	VAR with i	VAR without i	VAR with i	VAR without i
J-test	$\times$	$\surd$	$\times$	$\surd$	$\times$	$\times$
Stability	$\times$	$\surd$	$\times$	$\times$	$\times$	$\times$
Coeff. sign	$\times$	$\times$	$\surd$	$\surd$	$\surd$	$\surd$
Overall	$\times$	$\times$	$\times$	$\times$	$\times$	$\times$

Note:

$\surd$: 1. p value of the J-test or stability test is greater than 5%.

2. Estimated coefficients have the predicted signs.

$\times$: 1. p value of the J-test or stability test is less than 5%.

2. Estimated coefficients do not have the predicted signs.

3. In the overall evaluation, at least one of the three criteria fails and is marked $\times$.

Note that among the three investment equations, the only specification that seems to be fairly stable, and at the same time not rejected by the J-test, is the ΔI model. But the seemingly promising outcome is completely undermined by the incorrect signs on the estimated coefficients, (see the theoretical signs and values in Table 8 of the appendix for comparison). In addition, the observed stability in the ΔI model is conditional on the *exclusion* of the interest rate from the first-step model. This is an issue worth further scrutiny and is taken up in the next subsection.

A word of caution is in order about the interpretation of the estimation results. By construction, the J-test for cross-equation restrictions assesses the adequacy of the structural model by measuring the "distance" between the structural model and the corresponding reduced-form representation of the data, assuming a correctly specified reduced-form model. A restatement of this interpretation is to say, the J-test is essentially a *joint* test of the reduced-form model and the structural model. Note that results reported in Table 3 to Table 5 for the three investment models are computed on the basis of three *different* VARs, though the only difference lies in the decision variable. Though it is widely agreed that a carefully specified reduced-form VAR is able to well summarize the correlations present in the data, it is still possible with the inclusion of different decision variables, some VARs capture relatively less amount of correlations in the data than the others. Therefore, the J-tests based on different benchmarks are not directly comparable and the rejections seen in the data might not solely attribute to the misspecification in the structural model.

Table 3: Investment Equation in ΔI

	VAR with interest rate		VAR without interest rate	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2
$E_t \Delta \widehat{I}_{t+1}$	- 0.53	- 0.13	- 0.38	- 0.11
	[0.02]	[0.04]	[0.06]	[0.03]
$E_t \Delta \widehat{I}_{t+2}$	1.22	1.55	1.18	1.15
	[0.02]	[0.18]	[0.05]	[0.08]
$E_t \Delta \widehat{p}^I_{t+1}$	- 3.99	- 4.59	- 2.89	- 4.90
	[0.08]	[0.07]	[0.19]	[0.46]
$E_t \Delta \widehat{p}^I_{t+2}$	3.04	7.39	3.23	6.02
	[0.11]	[0.10]	[0.35]	[1.05]
$E_t \widehat{KY}_{t+1}$	- 0.52	- 0.40	- 0.39	- 0.34
	[0.01]	[0.08]	[0.02]	[0.09]
p(J-test)	0.00	0.04	0.08	0.84
p(QLR)	0.01	0.01	0.14	0.10
p(EW)	0.02	0.02	0.16	0.13
p(MW)	0.06	0.03	0.22	0.14

Note:

1. The investment equation under study is the third model in (36).
2. "Dataset 1" refers to the $\Delta I(NFB)$ dataset defined in the appendix. "Dataset 2" refers to the $\Delta I(GDP)$ dataset defined in the appendix.
3. For the estimated parameters, the numbers without brackets are the estimated coefficients. The numbers in the brackets are the standard deviations of the estimates.
4. p(J-test) is the p value of the test for over-identifying restrictions. p(QLR), p(EW) and p(MW) are p values of QLR, EW and MW tests.

Table 4: Investment Equation in IK

	VAR with interest rate		VAR without interest rate	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2
$E_t\widehat{IK}_{t+1}$	1.21	1.14	0.91	0.91
	[0.01]	[0.01]	[0.01]	[0.01]
$E_t\Delta\widehat{p}^I_{t+1}$	2.62	1.81	1.44	0.93
	[0.03]	[0.03]	[0.07]	[0.05]
$E_t\widehat{KY}_{t+1}$	0.50	0.43	1.95	1.78
	[0.01]	[0.01]	[0.04]	[0.04]
p(J-test)	0.00	0.02	0.13	0.26
p(QLR)	0.00	0.00	0.00	0.00
p(EW)	0.00	0.00	0.00	0.03
p(MW)	0.00	0.00	0.00	0.01

Note:

1. The investment equation under study is the first model in (36).
2. "Dataset 1" refers to the $IK(NFB)$ dataset defined in the appendix.
   "Dataset 2" refers to the $IK(GDP)$ dataset defined in the appendix.
3. For the estimated parameters, the numbers without brackets are the estimated coefficients. The numbers in the brackets are the standard deviations of the estimates.
4. p(J-test) is the p value of the test for over-identifying restrictions.
   p(QLR), p(EW) and p(MW) are p values of QLR, EW and MW tests.

Table 5: Investment Equation in ΔK

	VAR with interest rate		VAR without interest rate	
	Dataset 1	Dataset 2	Dataset 1	Dataset 2
$E_t \Delta \widehat{K}_{t+1}$	1.28	1.12	0.98	0.92
	[0.02]	[0.03]	[0.04]	[0.03]
$E_t \Delta \widehat{p}^I_{t+1}$	0.13	0.03	0.09	0.05
	[0.01]	[0.03]	[0.04]	[0.02]
$E_t \widehat{KY}_{t+1}$	0.12	0.07	0.08	0.09
	[0.01]	[0.02]	[0.02]	[0.02]
p(J-test)	0.00	0.01	0.05	0.04
p(QLR)	0.00	0.01	0.00	0.00
p(EW)	0.00	0.01	0.03	0.00
p(MW)	0.00	0.01	0.02	0.01

Note:

1. The investment equation under study is the second model in (36).
2. "Dataset 1" refers to the $\Delta K(NFB)$ dataset defined in the appendix. "Dataset 2" refers to the $\Delta K(GDP)$ dataset defined in the appendix.
3. For the estimated parameters, the numbers without brackets are the estimated coefficients. The numbers in the brackets are the standard deviations of the estimates.
4. p(J-test) is the p value of the test for over-identifying restrictions. p(QLR), p(EW) and p(MW) are p values of QLR, EW and MW tests.

6.4.3 The effects of interest rates

One observation that seems to deserve special attention is the effects on the second-step results of including/excluding interest rates in the first-step model. From Table 2 to Table 5, it seems excluding interest rates from the first-step VAR improves the performance of the second-step model to varying degrees. For instance, for the IK model and the ΔK model, excluding interest rates makes the estimated coefficients on $E_t\widehat{IK}_{t+1}$ and $E_t\widehat{\Delta K}_{t+1}$ much close to their theoretical value of 0.99. Moreover, for the ΔI model and the IK model, more criteria are satisfied with such an exclusion. In the previous literature, when the IK model was estimated by GMM $-$ an asymptotically equivalent approach to the two-step method under certain conditions, according to Li (2004) $-$ a typical instrument set would not include interest rates. It follows that the present exercise appears to suggest the supporting results reported in the literature might be sensitive to the choice of instruments, in particular, the interest rates.

In order to see if my concern above is well-founded, I investigate the effects of interest rates in the first-step model. Take the ΔI model as an example. Tables 13 and 14 in the appendix presents the stability tests and the Granger causality test of two VAR specifications: one includes interest rate and the other does not. From Table 13, it is evident interest rate is an important source of reduced-form instability. With the exclusion of the interest rate, the estimated value of λ, which is proportional to the standard deviation of the period-to-period change in the reduced-form coefficients, decreases by 14% to 34% for for the VAR equations. This may partially explain the improved stability in the ΔI model after interest rate is excluded. From Table 14, on the other hand, it seems that the ability of p^I, the price of capital stock, in explaining the behavior of other VAR variables depends on whether interest rate is included. When interest rate is excluded from the VAR specification, p^I helps explain most of the VAR variables. On the other hand, when both the price of capital and the interest rate show up in the VAR, p^I appears to lose much of its explanatory power on other variables except for p^I itself.

The sensitivity of the empirical results to interest rates naturally raises the question of whether interest rates should be included in the reduced-form model. On the one hand, the exclusion of interest rates might be justified on the ground that, interest rates affects investment decision through their role in determining the purchase price of capital goods. In other words, the explanatory power of interest rates on other variables might have been implicitly included in the purchase price of capital goods. As a consequence, in making their optimal investment decisions, it might be redundant for private firms to include both purchase price of capital and interest rate in their forecasting system. However, the above argument might only make sense in a partial equilibrium setup. It seems much less convincing if one thinks about the first-step model as a reduced-form representation of some general equilibrium model. There is an extensive literature on dynamic stochastic general equilibrium models, in particular, that on monetary business cycle models, which emphasizes the link between central bank's interest rate policies and investment[69]. Along this line of thinking, it would be a severe drawback not to incorporate interest rates to investment forecasting and decision making.

6.5 Further discussions

Note that in the previous investigation, regardless of interest rate being included in the first-step VAR or not, it appears hard for the standard investment models to account for the data correlations captured in the reduced-form model. Various factors could be responsible for this empirical outcome. First, inaccurate data could always be one source of the statistical failure. Difficulties in rigorously aggregating across heterogeneous firms and heterogeneous types of capital make structural analysis for aggregate investment a perilous task. In addition, the construction of the quarterly capital series through interpolation[70] would unavoidably introduce measurement error that may further deteriorate the data problem.

Second, the Euler equations estimated may just be incorrectly specified models. Although there has seemed to be very little skepticism about the modeling potential of these standard investment Eulers to macroeconomics, the tight restrictions imposed on the dynamic structure by these standard models appear not easy to fit the U.S. data. Therefore, a hard question might be to find a tractable Euler equation for investment that is not mis-specified.

It is useful to point out, statistical tests typically check whether data contain correlations that are not predicted by the model under test. In reduced-form models, such as vector autoregressions, it is relatively straightforward to amend the model to take into account the correlations present in the data. It is much more difficult to do so in Euler equations because the estimated parameters come from private agents' objective functions. When correlations present in the data are not predicted by the objective functions under study, it is often difficult to know how to parsimoniously change the objective functions. So it should not be too surprising the standard investment Eulers are statistically rejected by the data. Since these models are extremely simple in many dimensions. It is, perhaps, more important to discuss whether these statistical rejections have economic meaning, which we can learn from.

To improve the empirical performance of investment models, a variety of extensions or revisions might be worth considering. Here I just discuss a few. First, in models introduced so far, firm's time discount factor, β, capital share in production, α, and the depreciation rate

[69] Most recent work on this issue includes Dupor (2001 and 2002), Carlstrom and Fuerst (1999 and 2000), Li (2003) and Woodford (2003).

[70] The empirical exercise of this section is based on quarterly data while the quarterly series of capital stock is not directly available. Following the usual practice, I interpolate a quarterly series from the annual series of capital stock. See the data appendix for more details about this interpolation. In this regard, the model in ΔI, whose decision variable, the change of investment, is directly available on quarterly basis, is less affected by the measurement error caused by interpolation.

of capital, δ, are all assumed constant over the sample period. This might not be true under more realistic situations. For instance, uncertainty about discount factors is widely discussed in the theoretical macro literature. The argument is that firms are uncertain not only about what their future profits will be, but also about how those profits will be valued. The capital share, arguably would evolve over time following the progress in technology. The depreciation rate may also follow certain time variations exogenous to the optimal decision in investment. Appropriately accounting for these exogenous time variations help improve the model fit, say, by absorbing part of the instability observed in the current models [71].

Second, a re-evaluation of the proper modeling of the adjustment costs might be necessary. For instance, in the standard investment models, the adjustment costs of capital stock are specified as functions of investment and/or capital only. There is no channel to account for the effects from labor employment, the other production input, in the adjustment process. As a result, the optimality conditions in investment decision are independent of labor employment. A more careful modeling of investment may introduce employment to the cost function to allow for a simultaneous adjustment in both production factors. Another possibility might be to give up on the convex adjustment costs and introduce non-convexities, for instance, better models of investment might be provided by embedding irreversibilities, as suggested by a large literature, see, for example, Bernanke (1983), Abel and Eberly (1994) and Dixit and Pindyak (1994).

7 Concluding Remarks

Lucas' (1976) paper has strongly influenced the course of the economics profession for nearly three decades. As a result, Euler equation modeling strategy has become popular in both theoretical and empirical work. But there has been surprisingly little examination of the structural stability of the Euler equations − arguably one of the most relevant criterion for model validation by the standards of the Lucas critique. Thus, one purpose of this paper has been to formalize the necessity of structural stability testing. This is done by formulating the basic economic rationale of the Lucas critique in the classical two-step minimum distance framework: the time-varying reduced-form model in the first step reflects private agents' adaptation of their forecasts and behavior to the changing economic environment; and the presumed ability of Euler equations to deliver stable parameters indexing tastes and technology is interpreted as a time-invariant second-step model. Within this setup, I have been able to show, complementary to and independent of one another, both standard specification tests and stability assessment are needed for the evaluation of Euler equations. Therefore, the conventional model validation

[71] Oliner, Rudebusch and Sichel (1996) assume a time-varying discount factor and a constant depreciation rate of capital in their GMM estimation of a model in IK. According to their results, relaxing only the constancy of the discount factor seems not enough to induce structural stability.

strategy is a one-sided strategy by neglecting the stability assessment.

A second purpose of this article has been to describe how estimation and testing may be carried out in unstable econometric models when restrictions between unstable coefficients exist at each point in time. The approach adopted in the paper has demonstrated that, although estimation of such models by standard methods may induce invalid inference in general, testing for such restrictions raise no new technical issues since it only requires the standard testing procedure be supplemented by standard stability tests. However, a word of caution is required in interpreting the above statement. When there are other data problems that co-exists with instability, the proposed testing procedure could cease to work. For instance, it can be shown if instability is accompanied by highly persistent series, the compound effects will invalidate the proposed testing procedure.

Another important econometric issue not addressed in the application of this paper is weak identification. Weak identification can be a concern in the estimation of macroeconomic equations with expectations, as pointed out by Ma (2002), Mavroeidis (2001) and Fuhrer and Rudebusch (2002). Even if simulation investigation suggests the ML estimator (which is equivalent to the two-step estimator, as I have briefly mentioned in Section 5) less suffer from this problem than the GMM estimator, in principle weak identification might still arise if the second-step objective function is, say, flat around the minimum or locally non-quadratic. It has been shown that, if weak identification exists, the sampling distribution of conventional estimators are in general non-normal, and hence point estimates and inference based on standard methods are unreliable. (See Staiger and Stock (1997), Stock, Wright and Yogo (2002) and Stock and Yogo (2003a, 2003b) on the issue of weak instruments.) Note that in the context of a time-varying coefficient model, the presence of weak identification creates a problem involving *both* unstable coefficients *and* weak identification. Although it is possible to show the distortion on inference is dominated by weak identification, as illustrated in the appendix in a pilot study for the unstable 2SLS models, the consequence on estimation, inference and testing is not thoroughly understood. For instance, for practical purpose, whether tools developed to detect weak identification in a stable environment directly apply to models with time-varying coefficients is unclear.

8 References

Abel, A. (1980), "Empirical Investment Equations: An Integrative Framework, *Carnegie-Rochester conference series on public policy*, 12, 39 91.

Abel A. and O. Blanchard, (1983), "An Inter-temporal Equilibrium Model of Saving and Investment, *Econometrica* 51, 575 692.

Abel A. and J. Eberly (1994), "A Unified Model of Investment under Uncertainty", *American Economic Review*, 84, 1369 - 1384.

Andrews, D. and W. Ploberger, (1992) "Generic Uniform Convergence", *Econometric Theory*, 8, 241 - 257.

Andrews, D. and W. Ploberger, (1994) "Optimal Tests When a Nuisance Parameter is Present Only Under the Alternative", *Econometrica* 62, 1383 - 1414.

Altonji, G. and L. Segal (1996), "Small Sample Bias in GMM Estimation of Covariance Structure", *Journal of Business and Economic Statistics*, 14, 353 - 366.

Baba, Y., D. Hendry and R. Starr (1992), "The demand for M1 in the U.S.A, 1960 - 1988", *Review of Economic Studies*, 59, 25 - 61.

Baxter, M. and M. Crucini (1993) "Explaining Saving-Investment Correlations", *American Economic Review"*, 83, 416 - 436.

Bernanke, B. (1983), "Irreversibility, Uncertainty, and Cyclical Investment". *Quarterly Journal of Economics*, 98, 85 - 106.

Bernanke, B., M. Gertler and S. Gilchrist, (1998), "The financial accelerator in a quantitative business cycle framework, NBER working paper 6455.

Bernanke, B., M. Gertler and M. Watson (1997), "Systematic Monetary Policy and the Effects of Oil Price Shocks", *Brookings Papers on Economic Activity* 1, 91 - 142.

Blanchard, O. and S. Fischer, (1989), *Lectures on Macroeconomics.* MIT Press.

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

Boivin, J. and M. Watson (1999), "Time-Varying Parameter Estimation in a Linear IV Framework", mimeo, Columbia University.

Campbell J. and R. Shiller (1987), "Cointegration and Tests of Present Value Models", *Journal of Political Economy* 95 1062 - 1088.

Carlstrom, C. and T. Fuerst (2000), "The Role of Investment Spending in Sticky Price Models", mimeo, Federal Reserve Bank of Cleveland.

Carlstrom, C. and T. Fuerst (2000), "Forward-looking versus Backward-looking Taylor Rules", mimeo, Federal Reserve Bank of Cleveland.

Casares, M. and B. McCallum, (2000), "An optimizing IS-LM framework with endogenous investment, NBER working paper 7908.

Chiang, C.L. (1956), "On Regular Best Asymptotically Normal Estimates", *Annals of Mathematical Statistics*, 27, 336 - 351.

Chow, G. (1960), "Tests of Equality Between Sets of Coefficients in Two Linear Regressions", *Econometrica*, 28, 591 - 605.

Christiano L., M. Eichenbaum and C. Evans (2001), "Nominal Rigidities and the Dynamic Effect of a Shock to Monetary Policy", NBER working paper 8403.

Davidson, J. (1994), *Stochastic Limit Theory*, Oxford University Press Inc.

Dixit, A. and R. Pindyck (1994), *Investment under Uncertainty* Princeton University Press.

Dupor, W. (2001), “Investment and Interest Rate Policy”, *Journal of Economic Theory* 98, 85 - 113.

Dupor, W. (2002), “Interest Rate Policy and Investment with Adjustment Cost”, mimeo.

Edge, R. (2000),“Time-to-build, time-to-plan, habit-persistence and the liquidity effect, mimeo. Federal Reserve Board.

Edge, R. , T. Laubach and J. Williams (2003), “The Response of Wages and Prices to Technology Shocks”, mimeo, Federal Reserve Board.

Elliott, G. and U. Mueller, (2003a), “Optimally Testing General Breaking Processes in Linear Time Series Models”, mimeo, Princeton University.

Elliott, G. and U. Mueller, (2003b), “Breaks in Parameter of Econometric Models”, Research proposal, Princeton University.

Engle, R., D. Hendry and J. Richard (1983), “Exogeneity”, *Econometrica*, 51, 277 - 304.

Li, H. and U. Mueller (2004), “Valid Inference in Generalized Method of Moments Models with Time-Varying Coefficients”.

Engle, R. and D. Hendry (1993), “Testing super exogeneity and invariance in regression models”, *Journal of Econometrics*, 56, 119 - 139.

Ericsson, N. (1992), “Parameter constancy, mean square forecast errors, and measuring forecast performance: An exposition, extensions, and illustration”, *Journal of Policy Modeling*, 14, 465 - 495.

Ericsson, N. and J. Irons (eds.) (1994), *Testing Exogeneity*, Oxford, Oxford University Press.

Favero, C. and D. Hendry (1992), “Testing the Lucas Critique: A Review”, Econometric Reviews, 11, 265 - 306.

Ferguson, T.S. (1958), “A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities”, *Annals of Mathematical Statistics*, 29, 1046 - 1062.

Fuhrer, J. and G. Moore (1995), “Inflation Persistence”, *Quarterly Journal of Economics* 110, 127 - 160.

Fuhrer, J., G. Moore and S. Schuh (1995), “Estimating the Linear Quadratic Inventory Model: Maximum Likelihood versus Generalized Method of Moments”, *Journal of Monetary Economics* 35, 115 - 157.

Fuhrer, J. and G. Rudebusch (2002), “Estimating the Euler Equation for Output”, mimeo, Federal Reserve Bank of Boston.

Gertler, M, G. Hubbard and A. Kashyap, (1991), “Interest rate spreads, credit constraints, and investment fluctuations: An empirical investigation in *Financial markets and financial crises*, University of Chicago Press, 11- 31.

Ghysels, E. and A. Hall, (1990a), “Are consumption-based intertemporal capital asset pricing models structural?, *Journal of Econometrics* 45, 121-139.

Ghysels, E. and A. Hall, (1990b), “A test for structural stability of Euler conditions parameters estimated via the generalized method of moments estimator, *International Economic Review* 31, 355-364.

Gilchrist, S., (1990), “An empirical analysis of corporate investment and financing hierarchies using panel data, mimeo, Boston University.

Gilchrist, S. and C. Himmelberg, (1998), “Investment, fundamentals and finance, NBER working paper 6652.

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

Hansen, L., J. Heaton, and A. Yaron (1996), "Finite Sample Properties of Some Alternative GMM estimators", *Journal of Business and Economic Statistics*, 14, 262 - 280.

Hayashi, F. (1982), "Tobins marginal and average q: A neoclassical interpretation, *Econometrica* 50, 213 224.

Hayashi, F. (2000), *Econometrics*, Princeton University Press.

Hendry, D. (1988), "The Encompassing Implications of Feedback versus Feedforward Mechanisms in Econometrics", *Oxford Economic Papers*, 40, 132 - 149.

Hendry, D. and N. Ericsson (1991a), "An econometric analysis of U.K. money demand", *American Economic Review*, 81, 8 - 38.

Hendry, D. and N. Ericsson (1991a), "Modeling the Demand for Narrow Money in the U.K. and the U.S.", *European Economic Review*, 35, 833 - 881.

Hendry, D. and J. Richard (1982) "On the Formulation of Empirical Models in Dynamic Econometrics", *Journal of Econometrics*, 20, 3 - 33.

Hubbard, R. and A. Kashyap (1992), "Internal Net Worth and the Investment Process: An application to U.S. Agriculture", *Journal of Political Economy*, 100, 506 - 534.

Jermann, U. (1998), "Asset pricing in production economies, *Journal of Monetary economics* 41, 257 275.

Jondeau, E. and H. LeBihan (2001), "Testing for a Forward-Looking Phillips Curve: Additional Evidence from European and US Data", manuscript, Bank of France.

Kim, J. (1999) "Indeterminacy and Investment Adjustment Costs", mimeo, Federal Reserve Board.

King, R. and A. Kurmann (2002), "Expectations and the Term Structure of Interest Rates: Evidence and Implications", Federal Reserve Bank of Richmond *Economic Quarterly*.

Kurmann, A. (2003), "Maximum Likelihood Estimation of Dynamic Stochastic Theories with an Application to New Keynesian Pricing", manuscript, University of Quebec.

King, R. and M. Watson (1996), "Money, Prices, Interest Rates and the Business Cycle", *The Review of Economics and Statistics*, 78, 35 - 53.

Lettau, M. and S. Ludvigson (2001), "Consumption, Aggregate Wealth and Expected Stock Returns", *Journal of Finance*, 56, 815 - 849.

Leybourne, S. (1993), "Estimation and Testing of Time-varying Coefficient Regression Models in the Presence of Linear Restrictions", *Journal of Forecasting*, 12, 49 - 62.

Li, H. (2003), "Inflation Determination under a Taylor Rule: Consequences of Endogenous Capital Accumulation", mimeo, Princeton University.

Li, H. (2004), "Testing Alternative Models of Aggregate Supply", mimeo, Princeton University.

Lucas, R. (1976), "Econometric Policy Evaluation: A Critique", *The Phillips Curve and Labor Markets*, eds. K. Brunner and A. Melzer, Carnegie-Rochester Conference Series on Public Policy Vol. 1. 19 - 46.

Ma, A (2002) "GMM Estimation of the New Phillips Curve", *Economic Letters* 76, 411 - 417.

Mavroeidis, S. (2001), "Identification and Misspecification Issues in Forward Looking Monetary Models", mimeo, Oxford University.

Mueller, U. (2004), "A Theory of Robust Long-run Variance Estimation", mimeo, Princeton University.

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

Newey, W. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing", *Handbook of Econometrics*, eds. R. Engle and D. McFadden, Amsterdam: North-Holland, Volume 4, 2111 - 2245.

Nyblom, J. (1989), "Testing for the Constancy of Parameters Over Time", *Journal of American Statistical Association*, 78, 856 - 864.

Obstfeld M. and K. Rogoff, (1996), *Foundations of international macroeconomics*, MIT Press, MA.

Oliner, S., G. Rudebusch and D. Sichel, (1995), "New and old models of business investment: A comparison of forecasting performance, *Journal of Money, Credit and Banking* 27, 806-826.

Oliner, S., G. Rudebusch and D. Sichel (1996), "The Lucas Critique Revisited: Assessing the Stability of Empirical Euler Equations for Investment", *Journal of Econometrics* 70, 291 - 316.

Pindyck, R. and J. Rotemberg, (1983), "Dynamic factor demands under rational expectations, *Scandinavian Journal of Economics* 85, 223 238.

Pollard, D. (2001), *Contiguity*, unpublished draft.

Quandt, R. (1958), "The Estimation of the Parameters of a Linear Regression System Obeying Two Separate Regimes", *Journal of the American Statistical Association*, 53, 873 - 880.

Quandt, R. (1960), "Tests of the Hypothesis that a Linear Regression Obeys Two Separate Regimes", *Journal of the American Statistical Association* 55, 324- 330.

Resnick, S. (2001), *A Probability Path*, Birkhauser Boston.

Romer, D. (1996), *Advanced Macroeconomics*, McGraw-Hill Companies, Inc.

Rossi, B. (2004), "Optimal Tests for Nested Model Selection with Underlying Parameter Instability", mimeo, Duke University.

Saikkonene, P., and R. Luukonen, (1993), "Point Optimal Test for Testing the Order of Differencing in ARIMA Model", *Econometric Theory,* 9, 343 - 362.

Sargent, T (1979), "A Note on Maximum Likelihood Estimation of the Rational Expectations Models of the Term Structure", *Journal of Monetary Economics* 5.

Sbordone, A. (2003b), "A Limited Information Approach to the Simultaneous Estimation of Wage and Price Dynamics", mimeo, Rutgers University.

Shapiro, M. (1986a), "The dynamic demand for capital of labor, *Quarterly Journal of Economics* 101, 513 542.

Shapiro, M. (1986b), "Capital utilization and capital accumulation: Theory and evidence, *Journal of Applied Econometrics* 1, 211-234.

Shively, T. (1988), "An Exact Test of a Stochastic Coefficient i a Time Series Regression Model", *Journal of Time Series Analysis*, 9, 81 - 88.

Sowell, F. (1996), "Optimal Tests for Parameter Instability in the Generalized Method of Moments Framework ", *Econometrica*, 64, 1085 - 1108.

Staiger D., J. Stock and M. Watson (1997a), "The NAIRU, Unemployment and Monetary Policy", *Journal of Economics Perspectives* 11, 33 - 49.

Staiger D., J. Stock and M. Watson (1997b), "How Precise are Estimates of the Natural Rate of Unemployment", *Reducing Inflation: Motivation and Strategy* edited by C. Romer and D. Romer, University of Chicago Press.

Stock, J. (1991), "Confidence Intervals for the Largest Autoregressive Root in U.S. Economic Time

Series", *Journal of Monetary Economics*28, 435 - 460.

Stock, J. and D. Staiger (1997), "Instrumental Variables Regression with Weak Instruments", *Econometrica*, 65, 557 - 586.

Stock, J. and M. Watson (1996), "Evidence on Structural Instability in Macroeconomic Time Series Relations", *Journal of Business and Economic Statistics* 14, 11 - 29.

Stock, J. and M. Watson (1998), "Median Unbiased Estimation of Coefficient Variance in a Time-Varying Parameter Model", *Journal of American Statistical Association*, 93, 349 - 358.

Stock, J., J. Wright and M. Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments", *Journal of Business and Economic Statistics*, 20, 518 - 529.

Stock, J. and M. Yogo (2003), "Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments", mimeo, Harvard University.

Stock, J. and M. Yogo (2003), "Testing for Weak Instruments in Linear IV Regression" mimeo, Harvard University.

Strasser, H. (1985), *Mathematical Theory of Statistic: Statistical Experiments and Asymptotic Decision Theory*, New York: de Gruyter.

Smets, F. and R. Wouters (2003a), "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area", *European Economic Review*, forthcoming.

Smets, F. and R. Wouters (2003b), "Sources of Business Cycle Fluctuations in the U.S.: A Bayesian DSGE Approach", mimeo.

Timmermann, A. and B. Paye (2004), "Instability of Return Prediction Models", mimeo, UCSD.

Vaart, V. (1998), *Asymptotic Statistics*, Cambridge University Press. Cambridge, U.K.

Whited, T., (1992), "Debt, liquidity constraints and corporate investment: Evidence from panel data, *Journal of Finance* 47, 1425 1460.

Woodford, M., (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy* , Princeton University Press.

9 Appendix

9.1 Contiguity (Section 2.3)

Following the notation introduced in Section 2.3 for Assumption 4, let $\{z_t, t = 1, ..., T\}$ denote the data set when the sample size is T. ϕ_t is a $k \times 1$ vector of time-varying parameters in the likelihood function which follows the process defined in (9) and (10). The likelihood function of the data for a given ϕ is $f_T(z_1, ...z_T; \phi)$. In many cases, it can be written as a product of two terms, one that depends on ϕ, and another that does not: $f_T(z_1, ...z_T; \phi) = \prod_{t=1}^{T} f(z_1, ..., z_T; \phi) b(z_1, ...z_T)$ where $b(z_1, ...z_T)$ is the conditional distribution of some weakly exogenous variables at time t given all the preceding variables. To save notation, let $f_t(\phi) = f(z_1, ..., z_T; \phi)$ and $b_t = b(z_1, ...z_T)$. Then the density of the data under ϕ_0 is $\prod_{t=1}^{T} f_t(\phi_0) b_t$. The conditional density of the data under ϕ_t for a given ϕ path is $\prod_{t=1}^{T} f_t(\phi_t) b_t$. Accordingly, the unconditional density of the data under ϕ_t is $\int_{-\infty}^{+\infty} \prod_{t=1}^{T} f_t(\phi_t) b_t \; dv_\phi$ where v_ϕ is the measure of ϕ. Note that the ratio of the density under ϕ_t for a given path of ϕ and the density under ϕ_0 is provided by

$$LR_T = \prod_{t=1}^{T} f_t(\phi_t) b_t / \prod_{t=1}^{T} f_t(\phi_0) b_t = \exp\left\{ \sum_{t=1}^{T} \log f_t(\phi_t) - \sum_{t=1}^{T} \log f_t(\phi_0) \right\} \qquad (38)$$

Before introducing some regularity conditions of $f_t(\phi)$, define some notation. Let $I(\phi)$ denote $E(-\partial^2 \log f_t(\phi)/\partial\phi\partial\phi')$ under ϕ_0. Let $s_t(\phi)$ and $h_t(\phi)$ denote $\partial \log f_t(\phi)/\partial\phi$ and $\partial^2 \log f_t(\phi)/\partial\phi\partial\phi'$ respectively. Let $B(\phi, \delta)$ be an open ball in Φ_0 (where Φ_0 is the parameter space for ϕ) around ϕ with radius δ, ie $B(\phi, \delta) = \{\widetilde{\phi} \in \phi : \|\widetilde{\phi} - \phi\| < \delta\}$. Thus as δ approaches zero, $B(\phi, \delta)$ defines a shrinking neighborhood of ϕ. Define matrix $\overline{h}_t(\phi, \delta)$ with its element in ith row and jth column being $\overline{h}_t^{(i,j)}(\phi, \delta) = \sup\{h_t^{(i,j)}(\widetilde{\phi}) : \widetilde{\phi} \in B(\phi, \delta)\}$. Define $\underline{h}_t(\phi, \delta)$ likewise with its elements being $\underline{h}_t^{(i,j)}(\phi, \delta) = \inf\{h_t^{(i,j)}(\widetilde{\phi}) : \widetilde{\phi} \in B(\phi, \delta)\}$, for all i and j. Thus, the function $\underline{h}_t(\phi, \delta)$ and the function $\overline{h}_t(\phi, \delta)$ provide the lower bound and the upper bound of $h_t(\phi)$ on element-by-element basis within some neighborhood of ϕ. Finally, let $d(z_1, ..., z_t)$ be a dominating matrix with elements being defined as $d^{(i,j)}(z_1, ..., z_t) = \sup\{|h_t^{(i,j)}(\phi)| : \phi \in \Phi_0\}$. For notational simplicity, let $d_t = d(z_1, ..., z_t)$.

- Assumption A1: $\log f_t(\phi)$ *is continuously twice differentiable in* ϕ *for all* $\phi \in \Phi_0$.
- Assumption A2: $T^{-1/2} \sum_{t=1}^{[sT]} s_t(\phi_0) \Rightarrow \mathcal{N}(0,\; sI(\phi_0))$ *for all* $s \in [0, 1]$ *and* $I(\phi_0)$ *is positive definite.*
- Assumption A3: $\underline{h}_t(\phi, \delta)$ *and* $\overline{h}_t(\phi, \delta)$ *are uniform mixing of size* $-r/(2r-2)$ *or strong mixing of size* $-r/(r-2)$ *with* $r > 2$ *under* ϕ_0 *for any* t*, any* $\phi \in \Phi_0$ *and any* δ *sufficiently small. Moreover,* $\underline{h}_t(\phi, \delta)$, $\overline{h}_t(\phi, \delta)$ *and* d_t *are* L_r *bounded with* $r > 2$.
- Assumption A4: *For each* $\phi \in \Phi_0$*, there is a constant* $\tau > 0$ *such that* $\|\widetilde{\phi} - \phi\| \leq \tau$ *implies* $\|h_t(\widetilde{\phi}) - h_t(\phi)\| \leq F(z_1, ...z_t)\; g(\|\widetilde{\phi} - \phi\|)$ *for all* t*, where* $F(\cdot)$ *and* $g(\cdot)$ *are non-random functions with* $E(F(z_1, ..., z_t)) < \infty$ *and* $\lim_{y \to 0} g(y) = 0$.

Assumptions A1 and A2 are fairly standard in the literature. Assumption A3 places sufficient weak dependence condition on $\underline{h}_t(\phi, \delta)$ and $\overline{h}_t(\phi, \delta)$. Assumption A4 is a smoothness condition on $h_t(\phi)$, as a function of ϕ. It describes the distance between $h_t(\phi)$'s (evaluating at different ϕ values) as a function of the distance between the ϕ values. A restatement of Assumption A4 is, as τ becomes arbitrarily small, so that $\widetilde{\phi}$ and ϕ are arbitrarily close, $h_t(\widetilde{\phi})$ and $h_t(\phi)$ would be arbitrarily close to one another.

Before establishing contiguity in Lemma A2, a useful result is given below which will be used in the proof of Lemma A2 to determine the order of magnitude of the remainder term in the quadratic approximation of the log likelihood function.

Lemma A1: *Let $\bar{\phi}_t$ be any intermediate point between ϕ_0 and ϕ_t. Suppose for any realization of $\{\phi_t\}$, $T^{1/2}(\phi_{[sT]} - \phi_0) \to f_\phi(s)$ for any $s \in [0,1]$, where $f_\phi(\cdot)$ is a $k \times 1$ non-random, non-zero vector function that satisfies* $\sup_{s\in[0\ 1]} \|f_\phi(s)\| \leq K < \infty$*, then* $\sum(\phi_t - \phi_0)' \left(h_t(\bar{\phi}_t) - Eh_t(\phi_0)\right)(\phi_t - \phi_0) \xrightarrow{p} 0$.

Proof: Lemma 1 is proved in several steps. As a preparation, step 1 to step 3 below establish $T^{-1/2}\sum(h_t(\phi_0) - Eh_t(\phi_0))(\phi_t - \phi_0) \xrightarrow{p} 0$. This result is then used in steps 4 to 6 to establish Lemma A1.

Step 1: Some implications of Assumptions A3 and A4 are summarized as follows.

(a) Given Assumptions A3 and A4, according to Andrews' (1987) Corollary 1 and McLeish's (1975) Theorem 2.10, $\underline{h}_t(\phi,\delta)$ and $\overline{h}_t(\phi,\delta)$ satisfy pointwise LLNs for all $\phi \in \Phi_0$ and δ sufficiently small, i.e., $T^{-1}\sum \underline{h}_t(\phi,\delta) \xrightarrow{p} E(\underline{h}_t(\phi,\delta))$ and $T^{-1}\sum \overline{h}_t(\phi,\delta) \xrightarrow{p} E(\overline{h}_t(\phi,\delta))$.

(b) Assumption 4 implies $\lim_{\delta\to 0} E\underline{h}_t(\phi,\delta) = E(h_t(\phi))$ and $\lim_{\delta\to 0} E(\overline{h}_t(\phi,\delta)) = E(h_t(\phi))$, see Andrews' (1987) Corollary 2.

Step 2: $\bar{\phi}_t$ is in a shrinking neighborhood of ϕ_t because it is an intermediate point between ϕ_t and ϕ_0. This can be expressed as $\bar{\phi}_t \in B(\phi_0, \delta_T)$ where $\lim_{T\to\infty}\delta_T = 0$. Then, by the definition of $\underline{h}_t(\phi,\delta)$ and $\overline{h}_t(\phi,\delta)$, we have $\underline{h}_t(\phi_0,\delta_T) \leq h_t(\bar{\phi}_t) \leq \overline{h}_t(\phi_0,\delta_T)$ for all t and T, which implies

$$\sum(\phi_t-\phi_0)'(\underline{h}_t-Eh_t(\phi_0))(\phi_t-\phi_0) \leq \sum(\phi_t-\phi_0)'(h_t(\bar{\phi}_t)-Eh_t(\phi_0))(\phi_t-\phi_0) \leq \sum(\phi_t-\phi_0)'(\overline{h}_t-Eh_t(\phi_0))(\phi_t-\phi_0). \tag{39}$$

where $\underline{h}_t = \underline{h}_t(\phi_0,\delta_T)$ and $\overline{h}_t = \overline{h}_t(\phi_0,\delta_T)$. On the other hand, following result in part (b) of Step 1, $\text{plim}_{T\to\infty}\sum(\phi_t-\phi_0)'(\underline{h}_t(\phi_0,\delta_T) - E(\underline{h}_t(\phi_0,\delta_T)))(\phi_t-\phi_0) = \text{plim}_{T\to\infty}\sum(\phi_t-\phi_0)'(\underline{h}_t(\phi_0,\delta_T) - E(h_t(\phi_0)))(\phi_t - \phi_0)$, and likewise with $\underline{h}_t(\phi_0,\delta_T)$ being replaced by $\overline{h}_t(\phi_0,\delta_T)$. It follows that to show the convergence result in Lemma A1, it is sufficient to show $\sum(\phi_t-\phi_0)'(\underline{h}_t(\phi_0,\delta_T) - E(\underline{h}_t(\phi_0,\delta_T)))(\phi_t-\phi_0) \xrightarrow{p} 0$ and likewise with $\underline{h}_t(\phi_0,\delta_T)$ being replaced by $\overline{h}_t(\phi_0,\delta_T)$.

Step 3: To streamline notation, in what follows, $\underline{h}_t = \underline{h}_t(\phi_0,\delta_T)$ and $\overline{h}_t = \overline{h}_t(\phi_0,\delta_T)$. Let $\widetilde{y}_t^{ij} = (\underline{h}_t^{ij} - E\underline{h}_t^{ij}(\phi_0))(T^{1/2}(\phi_t^i - \phi_0^i))(T^{1/2}(\phi_t^j - \phi_0^j))$ where $\underline{h}_t^{ij}$ is the element of the ith row and the jth column in matrix $\underline{h}_t$. $Eh_t^{ij}(\phi_0)$ is defined likewise. ϕ_t^i is the ith element in vector ϕ_t, and ϕ_t^j, ϕ_0^i, ϕ_0^j are defined likewise. Then according to Assumption A3, $\underline{h}_t$ is L_r bounded, hence $\widetilde{y}_t^{ij}$ is also L_r bounded. To see this, by the L_r boundedness assumption, $\|\underline{h}_t^{ij}(\phi_0,\delta_T)\|_r \leq M < \infty$. Also, $\|E\underline{h}_t^{ij}(\phi_0,\delta_T)\|_r \leq M' < \infty$. Then,

$$\|\widetilde{y}_t^{ij}\|_r = \|(\underline{h}_t^{ij} - E\underline{h}_t^{ij})(T^{1/2}(\phi_t^i - \phi_0^i))(T^{1/2}(\phi_t^j - \phi_0^j))\|_r \leq \left(\|\underline{h}_t^{ij}\|_r + \|E\underline{h}_t^{ij}\|_r\right)(\sup_{s\in[0\ 1]}\|f_\phi(s)\|)^2 = [M+M']K^2 \tag{40}$$

which is independent of t. The inequality in (40) follows from the triangular inequality. Because $r > 2$, the derivation in (40) implies $\widetilde{y}_t$ is uniformly integrable.

Step 4: In what follows, we show that $\widetilde{y}_t$ is L_1 mixingale with respect to the σ-field $\mathcal{F}_t$ under both strong mixing and uniform mixing. Under strong mixing of size $-r/(r-2)$,

$$\left\|E(\widetilde{y}_t^{ij}|\mathcal{F}_{t-m})\right\|_1 \quad \leq \quad 6\ \alpha_m^{1-1/r}\ \|\widetilde{y}_t^{ij}\|_r \ \leq\ 6[M+M']K^2\ \alpha_m^{1-1/r}$$

where α_m is the mth strong mixing coefficient. Since $\alpha_m = O(m^{-r/(r-2)-\varepsilon})$ for some $\varepsilon > 0$, we have $\alpha_m^{1-1/r} = O(m^{-1-\varepsilon'})$ for $\varepsilon' = 1/(r-2) + ((r-1)/r)\varepsilon > 0$. Thus, under strong mixing, $\widetilde{y}_t$ is L_1 mixingale of size -1 with respect to constants that do not depend on t. Similarly, under uniform mixing,

$$\left\|E(\widetilde{y}_t^{ij}|\mathcal{F}_{t-m})\right\|_1 \quad \leq \quad 2\ \phi_m^{1-1/r}\|\widetilde{y}_t^{ij}\|_r \ \leq\ 2[M+M']K^2\ \phi_m^{1-1/r}$$

where ϕ_m is the mth uniform mixing coefficient. Since $\phi_m = O(m^{-r/(2r-2)-\varepsilon})$ for some $\varepsilon > 0$, we have $\phi_m^{1-1/r} = O(m^{-1/2-\varepsilon'})$ for $\varepsilon' = ((r-1)/r)\varepsilon > 0$. Thus, under strong mixing, $\widetilde{y}_t$ is L_1 mixingale of size $-1/2$

with respect to constants that do not depend on t.

Step 5: Theorem 19.11 of Davidson (1994) states that the sample mean of a uniformly integrable L_1 mixingale of any size with respect to constants that does not depend on t converges to zero in L_1-norm. So $T^{-1}\sum \widetilde{y}_t^{ij} \xrightarrow{L_1} 0$. But then convergence in L_1-norm implies convergence in probability, $T^{-1}\sum \widetilde{y}_t^{ij} \xrightarrow{p} 0$. This in turn gives $T^{-1}\sum_{t=1}^{T}\sum_{i=1}^{p}\sum_{j=1}^{p}\widetilde{y}_t^{ij} \xrightarrow{p} 0$, i.e., $\sum_{t=1}^{T}(\phi_t-\phi_0)'(\underline{h}_t - Eh_t(\phi_0))(\phi_t-\phi_0) \xrightarrow{p} 0$. Following the identical procedure, we obtain $\sum_{t=1}^{T}(\phi_t-\phi_0)'(\overline{h}_t - Eh_t(\phi_0))(\phi_t-\phi_0) \xrightarrow{p} 0$.

Then, sandwiched by two terms with probability limits of zero, the middle term in (39) also has a probability limit of zero, which is the convergence result in Lemma A1.

Lemma A2: *Suppose ϕ_t follows (9) and (10). Under Assumptions A1 to A3, the sequence of densities $\{\int_{-\infty}^{+\infty}\prod_{t=1}^{T} f_t(\phi_t)b_t dv_\phi : T \geq 1\}$ are contiguous to the sequence of densities $\{\prod_{t=1}^{T} f_t(\phi_0)b_t : T \geq 1\}$.*

Proof: We proceed in two steps. In the first step, the contiguity between the densities of the data under ϕ_0 and the conditional densities of the data under ϕ_t for a *given* path of ϕ is established. Let us call it "conditional" contiguity. In the second step, the "unconditional" contiguity stated in Lemma A2 is shown to hold given the "conditional" contiguity proved in the first step.

Step 1: In order to establish the "conditional" contiguity, the following result is used, which follows Lemma 9 of Pollard (2001)[72]: For a given path of $\phi_t, t = 1, ..., T$, if (i) LR_T defined in (38) converges weakly to some random variable LR under ϕ_0 and (ii) $E(LR) = 1$ under ϕ_0, then the conditional densities $\left\{\prod_{t=1}^{T} f_t(\phi_t)b_t : T \geq 1\right\}$ are contiguous to the densities $\left\{\prod_{t=1}^{T} f_t(\phi_0)b_t : T \geq 1\right\}$.

A quadratic expansion of $\sum \log f_t(\phi_t)$ around $\sum \log f_t(\phi_0)$ yields

$$\begin{aligned} & \sum \log f_t(\phi_t) - \sum \log f_t(\phi_0) \\ = & \sum s_t(\phi_0)'(\phi_t-\phi_0) + \frac{1}{2}\sum(\phi_t-\phi_0)'h_t(\bar{\phi}_t)(\phi_t-\phi_0) \\ = & \sum s_t(\phi_0)'(\phi_t-\phi_0) + \frac{1}{2}\sum(\phi_t-\phi_0)'Eh_t(\phi_0)(\phi_t-\phi_0) + R_T \end{aligned} \tag{41}$$

where $\bar{\phi}_t$ is an intermediate point between ϕ_0 and ϕ_t. The remainder term in (41) is $R_T = \frac{1}{2}\sum(\phi_t - \phi_0)'(h_t(\bar{\phi}_t) - Eh_t(\phi_0))(\phi_t-\phi_0) = o_p(1)$, which follows from the convergence result of Lemma A1. Moreover, it follows that

$$\exp\{R_T\} = 1 + \widetilde{R}_T \quad \text{with} \quad \widetilde{R}_T = o_p(1) \tag{42}$$

Next, the limit of (41) can be obtained by deriving the limit of the first order term and the limit of the second order term respectively. The limit of the first order term is

$$\begin{aligned} \sum s_t(\phi_0)'(\phi_t-\phi_0) &= \sum(T^{-1/2}(\phi_t-\phi_0))'\left[T^{-1/2}\sum_{i=1}^{t} s_i(\phi_0) - T^{-1/2}\sum_{i=1}^{t-1} s_i(\phi_0)\right] \\ &\Rightarrow \int f_\phi(r)'\, I(\phi_0)^{1/2}\, dW_s(r) \end{aligned} \tag{43}$$

where $W_s(\cdot)$ is a standard Weiner process associated with the score function. The limit in (43) follows from the uniform square integrability of each element of $f_\phi(s)$ for all $s \in [0,1]$ (which is an immediate

[72] The result is also described in, for example, Theorems 16.8 and 18.11 of Strasser (1985), Lemma 3 of Andrews and Ploberger (1992) and Lemma a-4 of Andrews and Ploberger (1994).

result of the boundness assumption of $f_\phi(s)$ as stated in Lemma A1). The limit of the second order term can be calculated as

$$\begin{aligned} \sum(\phi_t-\phi_0)'Eh_t(\phi_0)(\phi_t-\phi_0) &= T^{-1}\sum\left[T^{1/2}(\phi_t-\phi_0)'\right]Eh_t(\phi_0)\left[T^{1/2}(\phi_t-\phi_0)\right] \\ &\to -\int_0^1 f_\phi(r)'\,I(\phi_0)\,f_\phi(r)dr \end{aligned} \tag{44}$$

Thus the limit of the conditional likelihood ratio statistic can be computed as

$$\begin{aligned} LR_T &= \prod f_t(\phi_t)b_t/\prod f_t(\phi_0)b_t \\ &= \exp\left\{\sum s_t(\phi_0)'(\phi_t-\phi_0)+\frac{1}{2}\sum(\phi_t-\phi_0)'Eh_t(\phi_0)(\phi_t-\phi_0)+R_T\right\} \\ &= \exp\left\{\sum s_t(\phi_0)'(\phi_t-\phi_0)+\frac{1}{2}\sum(\phi_t-\phi_0)'Eh_t(\phi_0)(\phi_t-\phi_0)\right\}(1+\widetilde{R}_T) \\ &= \exp\left\{\sum s_t(\phi_0)'(\phi_t-\phi_0)+\frac{1}{2}\sum(\phi_t-\phi_0)'Eh_t(\phi_0)(\phi_t-\phi_0)\right\}+o_p(1) \\ &\Rightarrow \exp\left\{\int_0^1 f_\phi(r)'I(\phi_0)^{1/2}dW_s(r)-\frac{1}{2}\int f_\phi(r)'I(\phi_0)f_\phi(r)dr\right\} \end{aligned} \tag{45}$$

where the second equality follows from (41), the third equality follows from (42), the fourth equality is derived by using $\sum s_t(\phi_0)'(\phi_t-\phi_0)=O_p(1)$, $\sum(\phi_t-\phi_0)'Eh_t(\phi_0)(\phi_t-\phi_0)=O(1)$, and $(O_p(1)+O(1))o_p(1)=o_p(1)$. The final limiting distribution is obtained by using (43) and (44).

Note the limit of LR_T is a random variable whose randomness comes from $W_s(\cdot)$. Let $Z=\int f_\phi(r)'I(\phi_0)^{1/2}dW_s(r)-\frac{1}{2}\int f_\phi(r)'I(\phi_0)f_\phi(r)$. Then $LR_T\Rightarrow e^Z$. Since the only random component of the Z variable, $W_s(\cdot)$, is a Gaussian variable, Z is a Gaussian random variable whose distribution is given by $Z\sim N\left(-\frac{1}{2}\int f_\phi(r)'I(\phi_0)f_\phi(r)dr,\int f_\phi(r)'I(\phi_0)f_\phi(r)dr\right)$. Thus by the moment generating function of a normal variable,

$$E\left[e^Z\right]=\exp\left\{-\frac{1}{2}\int f_\phi(r)'I(\phi_0)f_\phi(r)dr+\frac{1}{2}\int f_\phi(r)'I(\phi_0)f_\phi(r)dr\right\}=1.$$

Step 2: Introduce some notation. Let $P_0(\mathcal{B}_T)$ denote the probability of event $\mathcal{B}_T$ under ϕ_0. Let $P_{1,\phi}(\mathcal{B}_T)$ denote the probability of event $\mathcal{B}_T$ under ϕ_t conditional on a given path of $\{\phi_t,t=1,...,T\}$. Let $P_1(\mathcal{B}_T)$ denote the unconditional probability of event $\mathcal{B}_T$ under ϕ_t. Then by definition, the contiguity result in step 1 for a given ϕ path can be stated as follows: *For any sequences of random variable $\mathcal{Y}_T$, $P_0(\mathcal{Y}_T\in\mathcal{A})\to 0$ for any set $\mathcal{A}$ implies $P_{1,\phi}(\mathcal{Y}_T\in\mathcal{A})\to 0$ for any set $\mathcal{A}$.*

To show contiguity under any given path of ϕ implies the "unconditional" contiguity under a random path of ϕ, we need to show: $P_{1,\phi}(\mathcal{Y}_T\in\mathcal{A})\to 0$ *for any set $\mathcal{A}$ implies* $P_1(\mathcal{Y}_T\in\mathcal{A})\to 0$ *for any set $\mathcal{A}$*. Note that $P_1(\mathcal{Y}_T\in\mathcal{A})=E(P_{1,\phi}(\mathcal{Y}_T\in\mathcal{A}))$ where the expectation is taken with respect to ϕ. To see this, let $f(\mathcal{Y}_T)$ be the unconditional density of $\mathcal{Y}_T$; $f(\mathcal{Y}_T|\phi)$ be the conditional density of $\mathcal{Y}_T$ for a given ϕ path; and $f(\phi)$ be the marginal density of ϕ. Let $1(\cdot)$ be the indicator function. Then,

$$\begin{aligned} P_1(\mathcal{Y}_T\in\mathcal{A}) &= \int 1_{(\mathcal{Y}_T\in\mathcal{A})}\,f(\mathcal{Y}_T)\,dv_{\mathcal{Y}_T} \;=\; \int 1_{(\mathcal{Y}_T\in\mathcal{A})}\left(\int f(\mathcal{Y}_T|\phi)\,f(\phi)\,dv_\phi\right)dv_{\mathcal{Y}_T} \\ &= \int\int 1_{(\mathcal{Y}_T\in\mathcal{A})}\,f(\mathcal{Y}_T|\phi)\,f(\phi)\,dv_\phi\,dv_{\mathcal{Y}_T} \;=\; \int\left(\int 1_{(\mathcal{Y}_T\in\mathcal{A})}\,f(\mathcal{Y}_T|\phi)\,dv_{\mathcal{Y}_T}\right)f(\phi)\,dv_\phi \\ &= \int P_{1,\phi}(\mathcal{Y}_T\in\mathcal{A})\,f(\phi)\,dv_\phi \;=\; E\left(P_{1,\phi}(\mathcal{Y}_T\in\mathcal{A})\right) \end{aligned}$$

where the fourth equality follows from the Fubini's theorem. Then, as a result of the Dominated Convergence Theorem (see Theorem 5.3.3 on pp.133 of Resnick (2001)), $E(P_{1,\phi}(\mathcal{Y}_T\in\mathcal{A}))\to 0$ because

(i) $P_{1,\phi}(\mathcal{Y}_T \in \mathcal{A}) \to 0$; (ii) $P_{1,\phi}(\mathcal{Y}_T \in \mathcal{A})$ is bounded by 1. Combining this result and the result from step 1, the sequence of densities of the data under ϕ_0 and the sequence of unconditional densities of the data under ϕ_t are contiguous.

9.2 Linearization of $g(\phi_t, \theta_0)$ (Section 3.1)

A local linearization of $g(\phi_t, \theta_0)$ around ϕ_0 leads to $g(\phi_t, \theta_0) = g(\phi_0, \theta_0) + D_g(\phi_t - \phi_0) + R_T$ where R_T is the remainder term. Let $G^{(i)}_{\phi\phi}(\bar{\phi}_t, \theta_0) = \partial^2 g^{(i)}(\bar{\phi}_t, \theta_0)/\partial\phi\partial\phi'$, then $R_T = \frac{1}{2}\sum_{i=1}^{l}(\phi_t - \phi_0)' G^{(i)}_{\phi\phi}(\bar{\phi}_t, \theta_0)(\phi_t - \phi_0) = \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\sum_{k=1}^{l}(\phi_t^j - \phi_0^j)' G^{(ijk)}_{\phi\phi}(\bar{\phi}_t, \theta_0)(\phi_t^k - \phi_0^k)$ where l is the dimension of g. $\bar{\phi}_t$ is an intermediate point between ϕ_0 and ϕ_t. $\bar{\phi}_t$ is in the neighborhood of ϕ_0 because ϕ_t is local to ϕ_0. Then by (9), (10) and Condition 2, there exists $M < \infty$ such that $|R_T| = \frac{1}{2}|\sum_{i=1}^{l}(\phi_t - \phi_0)' G^{(i)}_{\phi\phi}(\bar{\phi}_t, \theta_0)(\phi_t - \phi_0)| \leq \frac{1}{2}\sum_{i=1}^{l}\sum_{j=1}^{l}\sum_{k=1}^{l}|(\phi_t^j - \phi_0^j)'||G^{(ijk)}_{\phi\phi}\bar{\phi}_t, \theta_0)||\phi_t^k - \phi_0^k)| \leq T^{-1}\widetilde{M}Q_T(t/T) = O_p(T^{-1})$ where $\widetilde{M} = \lambda^2 l M |\Sigma^{(j)}|^{1/2}|\Sigma^{(k)}|^{1/2} = O(1)$ where p is the dimension of ϕ, $\Sigma^{(j)}$ and $\Sigma^{(k)}$ are the jth and kth rows of the matrix Σ. $Q_T(t/T) = \sum_j\sum_k |W_T^j(t/T)||W_T^k(t/T)| = O_p(1)$ where $W_T^j(t/T)$ and $W_T^k(t/T)$ are the finite sample approximations of the standard Brownian motions associated with ϕ_t^j and ϕ_t^k. Thus $g(\phi_t, \theta_0) = g(\phi_0, \theta_0) + D_g(\phi_t - \phi_0) + O_p(T^{-1})$..

9.3 Lemma 1 (Section 2.3)

A lemma of Stock and Watson (1998) is reproduced below, which will be used in proving Lemma 1.

Lemma A3: *Let x_t be a mean zero stationary vector stochastic process with fourth-order cumulants that satisfy $\sum_{r1,r2,r3=-\infty}^{\infty}|C(r1, r2, r3)| < \infty$. Let w_t be either a scalar nonrandom sequence or a random variable that is independent of x_t for which $sup_i sup_{t\geq 1} E|x_{it}|^4 < \infty$ and $sup_{t\geq 1} E|w_t|^4 < \infty$. Then $T^{-1}\sum_{t=1}^{[sT]} x_t w_t \xrightarrow{p} 0$ uniformly in s.*

Proof: See the appendix of Stock and Watson (1998).

In what follows, I first prove Condition 1 *under* ϕ_0, that is, to show Condition 1 holds in the case where the $\{w_{it}\}'$s are generated by the corresponding stable process. The regressors in this hypothetical model are assumed to satisfy Assumption 1 to 3. The strategy of the proof *under* ϕ_0 follows Stock and Watson (1998). Let W_i and W_{ϕ_i} be independent k_i-dimensional standard Brownian motions. An implication of Assumptions 1 to 3 is as follows. Let $\Gamma_{ij} = E(w_{it}w_{jt}')$ and let $\Gamma^i_{w\varepsilon} = \mathrm{Var}(w_{it}\varepsilon_t)$ for $i = 1, 2$ and $j = 1, 2$.

(i) $T^{-1}\sum_{t=1}^{[sT]} w_{it}w_{jt}' \xrightarrow{p} s\Gamma_{ij}$;
(ii) $\Gamma^{i\ -1/2}_{w\varepsilon}\, T^{-1/2}\sum_{t=1}^{[sT]} w_{it}\varepsilon_t \Rightarrow W_i(s)$;
(iii) $\Sigma_i^{-1/2}\, T^{-1/2}\sum_{t=1}^{[sT]}\nu_{it} \Rightarrow W_{\phi_i}(s)$.

Next, I use Lemma A3 to prove the third requirement in Condition 1. To be more concrete, consider the case of $i = 1$ and $j = 1$. Then, $T^{-1}\sum_{t=1}^{[sT]} w_{1t}w_{1t}' T^{-1/2}\sum_{i=1}^{t}\nu_{1i} = \xi_{1T}(s) + \xi_{2T}(s)$ where $\xi_{1T}(s) = T^{-1}\sum_{t=1}^{[sT]}\Gamma_{11}T^{-1/2}\sum_{i=1}^{t}\nu_{1i}$ and $\xi_{2T}(s) = T^{-1}\sum_{t=1}^{[sT]}[w_{1t}w_{1t}' - \Gamma_{11}]\,T^{-1/2}\sum_{i=1}^{t}\nu_{1i}$. Limits are obtained for these terms.

(a) For $\xi_{1T}(s)$, the limit under ϕ_0 follows from result (iii), $\xi_{1T}(s) \Rightarrow \Gamma_{11}\Sigma_\beta^{1/2}\int_0^s W_{\phi_1}(r)dr$.

(b) For $\xi_{2T}(s)$, $w_{1t}w_{1t}' - \Gamma_{11}$ is stationary, mean zero and with absolutely summable fourth-order cumulants under ϕ_0. Let $\nu_{1t} = \Sigma_\beta^{1/2}\eta_{1t}$, where $\eta_{1t} \sim i.i.d.(0, I_{k1})$. Notice that $T^{-1/2}\sum_{i=1}^{t}\nu_i$

has four finite moments[73]. Thus $\xi_{2T}(s)$ satisfies the conditions of Lemma A3 with $z_t = w_{1t}w'_{1t} - \Gamma_{11}$ and $w_t = T^{-1/2}\sum_{i=1}^{t}\nu_{1t}$. As a result, $\xi_{2T}(s) = o_p(1)$. Combining results of (a) and (b), $T^{-1}\sum_{t=1}^{[sT]} w_{1t}w'_{1t}T^{-1/2}\sum_{i=1}^{t}\eta_{1i} \Rightarrow \Gamma_{11}\Sigma_{\beta}^{1/2}\int_0^s W_{\phi_1}(r)dr$. Thus Condition 1 is established under ϕ_0. Finally, by Assumption 4, all the above convergence results also hold under ϕ_t.

9.4 Proposition 1 (Section 3.1)

The relevant model is $y_t = w'_t\phi_t + \varepsilon_t$. Most of the derivation is covered in Section 3.1. The only missing part is the derivation of the limit of A_{2T}. Let $x_{1t} = w_t$ and $\nu_{1t} = \nu_t$ in Condition 1. This leads to $T^{-1/2}\sum_{t=1}^{T} w_t w'_t(\phi_t - \phi_0) = \lambda T^{-1}\sum_{t=1}^{T} w_t w'_t T^{-1/2}\sum_{i=1}^{t}\nu_i \Rightarrow \lambda\ E(w_t w'_t)\ \Sigma^{1/2}\int W_{\phi}(r)dr$. By part 1 of Condition 1, $T^{-1}\sum w_t w'_t \xrightarrow{p} E(w_t w'_t)$ under ϕ_t. Therefore, the limit of A_{2T} under ϕ_t is $A_{2T} = \left[T^{-1}\sum w_t w'_t\right]^{-1} T^{-1/2}\sum w_t w'_t(\phi_t - \phi_0) \Rightarrow \lambda\Sigma^{1/2}\int W_{\phi}(r)dr$. Moreover, the a standard Brownian motion, $\int W_{\phi}(r)dr \Rightarrow N(0, 1/3)$. As a result, the limiting distribution of A_{2T} is given by $A_{2T} \Rightarrow N(0, \frac{1}{3}\lambda^2\Sigma)$.

9.5 Proposition 2 (Section 3.2)

The relevant model is $y_t = x'_t\beta_t + z'_t + \varepsilon_t$. Relying on the contiguity argument, in order to show Proposition 2 *under* $\{\beta_t, \gamma_t\}$, I only need to show it holds *under* $\{\beta_0, \gamma_0\}$, that is, as if the regrssors x_t and z_t are generated by the stable model.

To analyze in matrix form, define $T\times k$ matrix $X = [x_1, ..., x_T]'$, $T\times(k-d)$ matrix $Z = [z_1, \cdots, z_T]'$, $T\times d$ matrix $U = [u_1, \cdots, u_T]'$, $T\times 1$ vector $\varepsilon = [\varepsilon_1, \cdots, \varepsilon_T]'$. Let $(k-d)T\times 1$ vector $\widetilde{\gamma} = [(\gamma_1 - \gamma_0)', ..., (\gamma_T - \gamma_0)']'$ and $T\times(k-d)T$ matrix $\widetilde{Z} = diag(z'_1, ..., z'_T)$. Also define the $(kT)\times 1$ vector $\widetilde{\beta}$ and the $T\times(kT)$ matrix $\widetilde{X}$ likewise.

Model (16) can then be written in matrix form as $y = X\beta_0 + Z\gamma_0 + \widetilde{\varepsilon}$ with $\widetilde{\varepsilon} = \varepsilon + \widetilde{X}\widetilde{\beta} + \widetilde{Z}\widetilde{\gamma}$. where $\widetilde{X}\widetilde{\beta}$ and $\widetilde{Z}\widetilde{\gamma}$ represent the ignored instabilities from β_t and γ_t. The standard estimator of β_0 can be obtained by partialing out z_t. This is achieved by premultiplying the regression by the residual matrix $M_z = I - Z(Z'Z)^{-1}Z'$. Then the OLS estimator of β is solved as $\widehat{\beta} = (U'M_zU)^{-1}U'M_zy$. Accordingly,

$$T^{1/2}\ (\widehat{\beta} - \beta_0)\ =\ \left[T^{-1}U'M_zU\right]^{-1}\ T^{-1/2}U'M_z\widetilde{\varepsilon}. \tag{46}$$

It is straightforward to verify $T^{-1}U'M_zU \xrightarrow{p} E(u_t u'_t)$. So the distribution is driven by $T^{-1/2}U'M_z\widetilde{\varepsilon}$. Let $P_z = Z(Z'Z)^{-1}Z'$, then

$$T^{-1/2}\ U'M_z\widetilde{\varepsilon} = T^{-1/2}\ U'\widetilde{\varepsilon} - T^{-1/2}\ U'P_z\widetilde{\varepsilon}. \tag{47}$$

I want to show the behavior of $T^{-1/2}U'M_z\widetilde{\varepsilon}$ is (i) asymptotically independent of γ_t; (ii) but it depends on β_t. Thus, the distortion on the inference over β, if any, comes from the ignored time variation in β itself. The first term on the right-hand side of (47) is $T^{-1/2}U'\widetilde{\varepsilon} = T^{-1/2}U'\varepsilon + T^{-1/2}U'\widetilde{X}\widetilde{\beta} + T^{-1/2}U'\widetilde{Z}\widetilde{\gamma}$. The limits of the three right-hand side terms are (i) $T^{-1/2}U'\varepsilon \Rightarrow \mathcal{N}(0,\ \mathrm{Var}(u_t\varepsilon_t))$ This the standard result in the corresponding stable model. (ii) $T^{-1/2}U'\widetilde{X}\widetilde{\beta} = T^{-1/2}\sum_{t=1}^{T} u_t x'_t\left[\lambda_\beta\sum_{i=1}^{t}\nu_{1i}\right] \Rightarrow \lambda_\beta E(u_t u'_t)\Sigma_\beta^{1/2}\int W_\beta(r)dr$, which follows from part 3 of Condition 1 with $x_{1t} = u_t$ and $x_{2t} = x_t$ and uses the equality

[73] To see this, consider for notational simplicity $k_1 = 1$ (The argument for $k_1 > 1$ is similar), $\max_{t\in[1,T]} E\left[T^{-1/2}\sum_{i=1}^{t}\nu_i\right]^4 = T^{-2}\max_t\left[tE(\eta_i^4) + 3(t^2 - t)\left(E(\eta_i^2)\right)^2\right]\Sigma_\beta^2 = T^{-2}t^*\left[E(\eta_i^4) - 3\left(E(\eta_i^2)\right)^2\right]\Sigma_\beta^2 + 3T^{-2}t^{*2}\left[E(\eta_i^2)\right]^2\Sigma_\beta^2 \leq T^{-2}t^*\left[E(\eta_i^4) - 3\left(E(\eta_i^2)\right)^2\right]\Sigma_\beta^2 + 3\left[E(\eta_i^2)\right]^2\Sigma_\beta^2$. The first term above approaches zero in the limit, and the second term is finite.

$E(u_t x_t') = E(u_t u_t')$ in the partial regression of x_t on z_t. (iii) $T^{-1/2}U'\widetilde{Z}\widetilde{\gamma} = T^{-1/2}\sum_{t=1}^{T} u_t z_t' \left[\lambda_\gamma \sum_{i=1}^{t} \nu_{2i}\right] \Rightarrow \lambda_\beta E(u_t z_t')\Sigma_\gamma^{1/2} \int W_\gamma(r)dr = 0$ which follows from part 3 of Condition 1 with $x_{1t} = u_t$ and $x_{2t} = z_t$ and the zero limit is due to the orthogonal condition $E(u_t z_t') = 0$ in the partial regression. Note that (ii) reflects the effect of the time-varying β_t, and (iii) reflects the effect of the time-varying γ_t. It is evident that only β_t has non-trivial effect while γ_t's effect is *partialed out* asymptotically. Thus, (i) to (iii) above jointly yield

$$T^{-1/2}U'\widetilde{\varepsilon} \;\Rightarrow\; N(0,\ \mathrm{Var}(u_t\varepsilon_t)) + \lambda_\beta\ E(u_t u_t')\ \Sigma_\beta^{1/2} \int W_\beta(r)dr. \tag{48}$$

Turn to the second term of (47), which can be equivalently written as $T^{-1/2}U'P_z\widetilde{\varepsilon} = \left[T^{-1}U'Z\right]\left[T^{-1}Z'Z\right]^{-1}\left[T^{-1/2}Z'\widetilde{\varepsilon}\right]$. The limits of the the first two components *under* $\{\beta_0, \gamma_0\}$ are $T^{-1}U'Z \xrightarrow{p} E(u_t z_t') = 0$ and $T^{-1}Z'Z \xrightarrow{p} E(z_t z_t')$, following from part 1 of Condition 1, and the orthogonal condition in the partial regression of x on z. The derivation of the limit of the third component is similar to that of $T^{-1/2}U'\widetilde{\varepsilon}$, which results in $T^{-1/2}Z'\widetilde{\varepsilon} \Rightarrow N(0,\ \mathrm{Var}(z_t\varepsilon_t)) + \lambda_\beta\ E(z_t x_t')\ \Sigma_\beta^{1/2} \int W_\beta(r)dr + \lambda_\gamma\ E(z_t z_t')\ \Sigma_\gamma^{1/2} \int W_\gamma(r)dr.$. Taking together the limits of the three pieces, $T^{-1/2}U'P_z\widetilde{\varepsilon} \to 0.$. Then, together with (46), (47), (48), and let $V_\beta = E(u_t u_t')^{-1}\mathrm{Var}(u_t\varepsilon_t)E(u_t u_t')^{-1}$,

$$T^{-1/2}(\widehat{\beta} - \beta_0) \;\Rightarrow\; N(0,\ V_\beta) + \lambda_\beta\ \Sigma_\beta^{1/2} \int W_\beta(r)dr \;=\; N(0,\ V_\beta + \frac{1}{3}\lambda_\beta^2 \Sigma_\beta).$$

The limiting distribution of $\widehat{\gamma}$, the OLS estimator of γ_0, can be obtained likewise. Let $\Sigma_{uu} = E(u_t u_t')$ and $\Sigma_{zz} = E(z_t z_t')$, For future reference, I list the following results for $\widehat{\beta}$ and $\widehat{\gamma}$ where ϑ is the coefficient matrix in the partial regression of x_t on z_t.

$$T^{1/2}\ (\widehat{\beta} - \beta_0) \;=\; \Sigma_{uu}^{-1}\ T^{-1/2}U'\widetilde{\varepsilon} + o_p(1) \tag{49}$$

$$T^{1/2}\ (\widehat{\gamma} - \gamma_0) \;=\; \Sigma_{zz}^{-1}\ T^{-1/2}Z'\widetilde{\varepsilon} - \vartheta\ \Sigma_{uu}^{-1}\ T^{-1/2}U'\widetilde{\varepsilon} + o_p(1) \tag{50}$$

$$\widehat{\beta} - \beta_0 \;=\; O_p(T^{-1/2}) \quad \text{and} \quad \widehat{\gamma} - \gamma_0 \;=\; O_p(T^{-1/2}) \tag{51}$$

Discussion 1: The above derivation directly applies to the case that β is constant over time, by simply setting $\lambda_\beta = 0$. This gives $T^{-1/2}(\widehat{\beta} - \beta_0) \Rightarrow N(0,\ V_\beta)$. This simply says, the inference over the stable coefficients in an unstable model is asymptotically unaffected by the instability in the nuisance coefficients.

Discussion 2: To understand the insight behind the parallel result in Li and Muller (2004) that the inference over the stable coefficients in the *GMM model* is asympototically unaffected by the unstable nuisance coefficients, in what follows, I briefly describe the GMM model with parameter instability in Li and Muller (2004), and list the GMM first order condition through which, the analog between the unstable linear regression model and unstable GMM model is clarified. Thus, the technical insight in the linear regression models carries directly over to the non-linear GMM models.

The GMM moment condition in Li and Muller (2004) is $E(g(z_t; \beta_0, \gamma_t)) = 0$, where z_t is the observables, β_0 is the constant coefficients of main interest and γ_t is the unstable nuisance coefficients. The standard GMM estimator is defined as $[\widehat{\beta}, \widehat{\gamma}]' = \mathrm{argmin}_{\beta,\gamma} T^{-1}\sum g_t(\beta,\gamma)'W_T T^{-1}\sum g_t(\beta,\gamma)$ where W_T is some weighting matrix with $W_T \xrightarrow{p} W_0$. For regularity conditions imposed on the moment function and the data generating process, see Li and Muller (2004).

Define $\bar{g}_t(\beta,\gamma) = W_0^{1/2}g_t(\beta,\gamma)$, $\bar{G}_{\beta,T} = W_0^{1/2}T^{-1}\sum \partial g_t(\widehat{\beta},\widehat{\gamma})/\partial\beta$, $\bar{G}_{\gamma,T} = W_0^{1/2}T^{-1}\sum \partial g_t(\widehat{\beta},\widehat{\gamma})/\partial\gamma$, and $\bar{M}_{\gamma,T} = I - \bar{G}_{\gamma,T}[\bar{G}_{\gamma,T}'\bar{G}_{\gamma,T}]^{-1}\bar{G}_{\gamma,T}'$. Following the usual practice in econometrics that asymptotics are typically derived on the basis of the first-order approximation, a mean value expansion of the sample moment condition $T^{-1}\sum \bar{g}_t(\widehat{\beta},\widehat{\gamma})$ around $T^{-1}\sum \bar{g}_t(\beta_0,\gamma_t)$ yields

$$T^{-1}\sum \bar{g}_t(\widehat{\beta},\widehat{\gamma}) = [T^{-1}\sum \bar{g}_t(\beta_0,\gamma_t) + \bar{G}_{\gamma,T}T^{-1}\sum(\gamma_t-\gamma_0)] - \bar{G}_{\beta,T}(\widehat{\beta}-\beta_0) - \bar{G}_{\gamma,T}(\widehat{\gamma}-\gamma_0) + o_p(T^{-1/2}) \tag{52}$$

where the term $\bar{G}_{\gamma,T}T^{-1}\sum(\gamma_t-\gamma_0)$ captures the effects of the ignored time variation in γ_t. Note that in the linear regression model we have

$$e = (\varepsilon + \widetilde{Z}\widetilde{\gamma}) - X(\widehat{\beta}-\beta_0) - Z(\widehat{\gamma}-\gamma_0)$$

where e is the OLS residual matrix. Compare the two equations, the analog between the two models is obvious, and is summerized in Table 6. Moreover, the linearized GMM first order condition for β is

$$\begin{aligned}
o_p(T^{-1/2}) &= \bar{G}'_{\beta,T}\bar{M}_{\gamma,T}T^{-1}\sum \bar{g}_t(\widehat{\beta},\widehat{\gamma}) \\
&= \bar{G}'_{\beta,T}\bar{M}_{\gamma,T}T^{-1}\sum \bar{g}_t(\beta_0,\gamma_t) - \bar{G}'_{\beta,T}\bar{M}_{\gamma,T}\bar{G}_{\beta,T}(\widehat{\beta}-\beta_0) \\
&\quad -\bar{G}'_{\beta,T}\bar{M}_{\gamma,T}\bar{G}_{\gamma,T}(\widehat{\gamma}-\gamma_0) + \bar{G}'_{\beta,T}\bar{M}_{\gamma,T}\bar{G}_{\gamma,T}T^{-1}\sum(\gamma_t-\gamma_0) \\
&= \bar{G}'_{\beta,T}\bar{M}_{\gamma,T}T^{-1}\sum \bar{g}_t(\beta_0,\gamma_t) - \bar{G}'_{\beta,T}\bar{M}_{\gamma,T}\bar{G}_{\beta,T}(\widehat{\beta}-\beta_0)
\end{aligned}$$

where the third equality follows from $\bar{M}_{\gamma,T}\bar{G}_{\gamma,T} = 0$ which holds by construction. The above derivation completely resembles that in linear regressions. That is, when $\bar{M}_{\gamma,T}$ is used to partial out $\bar{G}_{\gamma,T}(\widehat{\gamma}-\gamma_0)$, the term $\bar{G}_{\gamma,T}T^{-1}\sum(\gamma_t-\gamma_0)$, representing the effects of the instability in γ_t, is also partialed out. Hence, asymptotic behavior of $\widehat{\beta}$ remains the same as that in the standard GMM model.

Table 6: Comparison between OLS and GMM

	OLS	GMM
Disturbance	ε	$\bar{\varepsilon}_T = \left[T^{-1}\sum \bar{g}_t(\beta_0,\gamma_t)\right]$
Residual	$e = y - X'\widehat{\beta} - Z'\widehat{\gamma}$	$\bar{e}_T = \left[T^{-1}\sum \bar{g}_t(\widehat{\beta},\widehat{\gamma})\right]$
Regressor 1	X	$\bar{G}_{\beta,T} = \left[T^{-1}\sum \partial\bar{g}_t(\beta,\gamma)/\partial\beta\right]$
Regressor 2	Z	$\bar{G}_{\gamma,T} = \left[T^{-1}\sum \partial\bar{g}_t(\beta,\gamma)/\partial\gamma\right]$
Residual matrix	$M_z = I - Z(Z'Z)^{-1}Z'$	$\bar{M}_{\gamma,T} = I - \bar{G}_{\gamma,T}(\bar{G}'_{\gamma,T}\bar{G}_{\gamma,T})^{-1}\bar{G}'_{\gamma,T}$
Normal equation	$X'M_ze = 0$	$\bar{G}'_{\beta,T}\bar{M}_{\gamma,T}\bar{e}_T = o_p(T^{-1/2})$

9.6 Proposition 3 (Section 4.1)

Recall that $J_T = Tg(\widehat{\phi},\widehat{\theta})'\widehat{W}g(\widehat{\phi},\widehat{\theta})$ where $\widehat{\phi}$ is the estimated first-step coefficients, $\widehat{\theta}$ is the efficient second-step estimator, and $\widehat{W} \xrightarrow{p} W_0$ with W_0 being the optimal weighting matrix and $\widehat{W}$ being a consistent estimator of W_0[74]. Use the notation introduced in Section 2.1, linearize $g(\widehat{\phi},\widehat{\theta})$ arond $g(\widehat{\phi},\theta_0)$,

$$g(\widehat{\phi},\widehat{\theta}) \;=\; g(\widehat{\phi},\theta_0)+G_\theta(\widehat{\phi},\theta_0)(\widehat{\theta}-\theta_0)+R_T \tag{53}$$

where $R_T=\frac{1}{2}\sum_i(\widehat{\theta}-\theta_0)'\partial^2 g^{(i)}(\widehat{\phi},\bar{\theta})/\partial\theta\partial\theta'(\widehat{\theta}-\theta_0)$ for $i\in[1,...l]$, where $\bar{\theta}$ is an intermediate point between $\widehat{\theta}$ and θ_0. Since (i) $\widehat{\theta}-\theta_0=O_p(T^{-1/2})$ and (ii) $g(\bar{\phi},\theta_0)^{(i)}/\partial\theta\partial\theta'$ is bounded according to Condition 2, $R_T=O_p(T^{-1})$ using an argument similar to that in Appendix 7.2. Standardize functions g and G_θ by the optimal weighting matrix and denote $\bar{g}(\phi,\theta)=W_0^{1/2}g(\phi,\theta)$, $\bar{G}_\theta(\phi,\theta)=W_0^{1/2}G_\theta(\phi,\theta)$. Then (53) can be written as $\bar{g}(\widehat{\phi},\widehat{\theta})=\bar{g}(\widehat{\phi},\theta_0)+\bar{G}_\theta(\phi_0,\theta_0)(\widehat{\theta}-\theta_0)+o_p(T^{-1/2})$ where I use $\bar{G}_\theta(\widehat{\phi},\theta_0)\xrightarrow{p}\bar{G}_\theta(\phi_0,\theta_0)$ which follows from the consistency of $\widehat{\phi}$ and the continuous mapping theorem. On the other hand, from the first order condition of the second-step problem, it is straightforward to derive $\widehat{\theta}-\theta_0=-\left[\bar{G}_\theta(\phi_0,\theta_0)'\bar{G}_\theta(\phi_0,\theta_0)\right]^{-1}\bar{G}_\theta(\phi_0,\theta_0)'\bar{g}(\widehat{\phi},\theta_0)+o_p(T^{-1/2})$. Substituting the expression into $\bar{g}(\widehat{\phi},\widehat{\theta})$ yields

$$\bar{g}(\widehat{\phi},\widehat{\theta}) \;=\; \bar{M}_\theta\;\bar{g}(\widehat{\phi},\theta_0)+o_p(T^{-1/2}) \tag{54}$$

where $\bar{M}_\theta=I-\bar{G}_\theta(\phi_0,\theta_0)[\bar{G}_\theta(\phi_0,\theta_0)'\bar{G}_\theta(\phi_0,\theta_0)]^{-1}\bar{G}_\theta(\phi_0,\theta_0)'$. $\bar{M}_\theta$ is an idempotent matrix with rank $T-l$ where $l=\dim(g)$. So $\bar{g}(\widehat{\phi},\widehat{\theta})$ is a functional of $\bar{g}(\widehat{\phi},\theta_0)$. As a result, J_T is a functional of $\bar{g}(\widehat{\phi},\theta_0)$ because $J_T=T\bar{g}(\widehat{\phi},\theta_0)'\bar{M}_\theta\bar{g}(\widehat{\phi},\theta_0)+o_p(1)$. According to the finding in Section 3.1, the distribution of $\bar{g}(\widehat{\phi},\theta_0)$ is unaffected by the time variation in ϕ_t. Thus, as a functional of $\bar{g}(\widehat{\phi},\theta_0)$, J_T is also asymptotically independent of ϕ_t. Moreover, note that $T^{1/2}\bar{g}(\widehat{\phi},\theta_0)\Rightarrow N(0,I)$. This in turn results in $J_T\Rightarrow\chi^2_{(l-r)}$, just as in the stable coefficient case.

9.7 Distributions of partial sums (Section 4.2.2)

The relevant model is $y_t=x_t'\beta_t+z_t'\gamma_t+\varepsilon_t$. Let $\widetilde{\varepsilon}_t=\varepsilon_t+x_t'(\beta_t-\beta_0)+z_t'(\gamma_t-\gamma_0)$. I derive the limiting distribution of $T^{-1/2}\sum_{t=1}^{[sT]}x_te_t$ in two steps.

1. Show $T^{-1/2}\sum_{t=1}^{[sT]}x_te_t$ can be expressed as a functional of $T^{-1/2}\sum_{t=1}^{[sT]}x_t\widetilde{\varepsilon}_t$.
2. Derive the distribution of $T^{-1/2}\sum_{t=1}^{[sT]}x_t\widetilde{\varepsilon}_t$ and hence the distribution of $T^{-1/2}\sum_{t=1}^{[sT]}x_te_t$.

Step 1: Let $\Sigma_{xx}=E(x_tx_t')$, $\Sigma_{xz}=E(x_tz_t')$, $\Sigma_{uu}=E(u_tu_t')$ and $\Sigma_{zz}=E(z_tz_t')$. Substitute $e_t=\widetilde{\varepsilon}_t-x_t'(\widehat{\beta}-\beta_0)-z_t'(\widehat{\gamma}-\gamma_0)$ into $T^{-1/2}\sum_{t=1}^{[sT]}x_te_t$, yields

$$T^{-1/2}\sum_{t=1}^{[sT]}x_te_t=T^{-1/2}\sum_{t=1}^{[sT]}\Big(x_t\widetilde{\varepsilon}_t-x_tx_t'(\widehat{\beta}-\beta_0)-x_tz_t'(\widehat{\gamma}-\gamma_0)\Big)$$

where the second right-hand side term can be decomposed in to $T^{-1/2}\sum_{t=1}^{[sT]}x_tx_t'(\widehat{\beta}-\beta_0)=A_{1T}(s)^\dagger+\Delta A_{1T}(s)$, with $A_{1T}(s)^\dagger=T^{-1}\sum_{t=1}^{[sT]}\Sigma_{xx}T^{1/2}(\widehat{\beta}-\beta_0)=s\Sigma_{xx}T^{1/2}(\widehat{\beta}-\beta_0)$, and $\Delta A_{1T}(s)=T^{-1}\sum_{t=1}^{[sT]}(x_tx_t'-\Sigma_{xx})T^{1/2}(\widehat{\beta}-\beta_0)=o_p(1)$, which follows from Condition 1(1) and (51). Similarly, $T^{-1/2}\sum_{t=1}^{[sT]}x_tz_t'(\widehat{\gamma}-\gamma_0)=$

[74] So long as $\widehat{W}$ is consistent of W_0 in the standard two-step model, $\widehat{W}$ remains consistent in the two-step TVP model by contiguity.

Table 7: Asymptotics of Partial Sums

$T^{-1/2}\sum_{t=1}^{[sT]} x_t e_t$	$= T^{-1/2}\sum_{t=1}^{[sT]} x_t\widetilde{\varepsilon}_t - sT^{-1/2}\sum_{t=1}^{T} x_t\widetilde{\varepsilon}_t + o_p(1)$
$T^{-1/2}\sum_{t=1}^{[sT]} x_t\widetilde{\varepsilon}_t$	$\Rightarrow \mathrm{Var}(x_t\varepsilon_t)^{1/2} g_\lambda(s)$
$T^{-1/2}\sum_{t=1}^{[sT]} x_t e_t$	$\Rightarrow \mathrm{Var}(x_t\varepsilon_t)^{1/2}(g_\lambda(s) - sg_\lambda(1))$
$g_\lambda(s)$	$= W(s) + \lambda_\beta\ D_{Bx}\int_0^s W_\beta(r)dr + \lambda_\gamma\ D_{Cx}\int_0^s W_\gamma(r)dr$
D_{Bx}	$= \mathrm{Var}(x_t\varepsilon_t)^{-1/2}\Sigma_{xx}\Sigma_\beta^{1/2}$
D_{Cx}	$= \mathrm{Var}(x_t\varepsilon_t)^{-1/2}\Sigma_{xz}\Sigma_\gamma^{1/2}$
$T^{-1/2}\sum_{t=1}^{[sT]} z_t e_t$	$= T^{-1/2}\sum_{t=1}^{[sT]} z_t\widetilde{\varepsilon}_t - sT^{-1/2}\sum_{t=1}^{T} z_t\widetilde{\varepsilon}_t + o_p(1)$
$T^{-1/2}\sum_{t=1}^{[sT]} z_t\widetilde{\varepsilon}_t$	$\Rightarrow \mathrm{Var}(z_t\varepsilon_t)^{1/2} f_\lambda(s)$
$T^{-1/2}\sum_{t=1}^{[sT]} z_t e_t$	$\Rightarrow \mathrm{Var}(z_t\varepsilon_t)^{1/2}(f_\lambda(s) - sf_\lambda(1))$
$f_\lambda(s)$	$= W(s) + \lambda_\beta\ D_{Bz}\int_0^s W_\beta(r)dr + \lambda_\gamma\ D_{Cz}\int_0^s W_\gamma(r)dr$
D_{Bz}	$= \mathrm{Var}(z_t\varepsilon_t)^{-1/2}\Sigma_{zx}\Sigma_\beta^{1/2}$
D_{Cz}	$= \mathrm{Var}(z_t\varepsilon_t)^{-1/2}\Sigma_{zz}\Sigma_\gamma^{1/2}$
$T^{-1/2}\sum_{t=1}^{[sT]} u_t e_t$	$= T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t - sT^{-1/2}\sum_{t=1}^{T} u_t\widetilde{\varepsilon}_t + o_p(1)$
$T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t$	$\Rightarrow \mathrm{Var}(u_t\varepsilon_t)^{1/2} h_\lambda(s)$
$T^{-1/2}\sum_{t=1}^{[sT]} u_t e_t$	$\Rightarrow \mathrm{Var}(u_t\varepsilon_t)^{1/2}(h_\lambda(s) - sh_\lambda(1))$
$h_\lambda(s)$	$= W(s) + \lambda_\beta\ D_{Bu}\int_0^s W_\beta(r)dr$
D_{Bu}	$= \mathrm{Var}(u_t\varepsilon_t)^{-1/2}\Sigma_{uu}\Sigma_\beta^{1/2}$

$s\Sigma_{xz}T^{1/2}(\widehat{\gamma}-\gamma_0)+o_p(1)$ Combining the above results,

$$\begin{aligned}
& s^{-1}\ T^{-1/2}\sum_{t=1}^{[sT]}\left[x_t x_t'(\widehat{\beta}-\beta_0)+x_t z_t'(\widehat{\gamma}-\gamma_0)\right] \\
= \quad & \Sigma_{xx}T^{1/2}(\widehat{\beta}-\beta_0)+\Sigma_{xz}T^{1/2}(\widehat{\gamma}-\gamma_0)+o_p(1) \\
= \quad & (\vartheta'\Sigma_{zz}\vartheta+\Sigma_{uu})\Sigma_{uu}^{-1}T^{-1/2}U'\widetilde{\varepsilon}+\vartheta'\Sigma_{zz}\left[\Sigma_{zz}^{-1}T^{-1/2}Z'\widetilde{\varepsilon}-\vartheta\Sigma_{uu}^{-1}T^{-1/2}U'\widetilde{\varepsilon}\right]+o_p(1) \\
= \quad & T^{-1/2}(U+Z\vartheta)'\widetilde{\varepsilon}+o_p(1)\ =\ T^{-1/2}X'\widetilde{\varepsilon}+o_p(1) \\
= \quad & T^{-1/2}\sum_{t=1}^{T}x_t\widetilde{\varepsilon}_t+o_p(1)
\end{aligned}$$

where I use (49), (50) and the results of partial regression that $\Sigma_{xx} = \vartheta'\Sigma_{zz}\vartheta + \Sigma_{uu}$ and $\Sigma_{xz} = \vartheta'\Sigma_{zz}$. Collecting terms, $T^{-1/2}\sum_{t=1}^{[sT]} x_t e_t = T^{-1/2}\sum_{t=1}^{[sT]} x_t\widetilde{\varepsilon}_t - sT^{-1/2}\sum_{t=1}^{T} x_t\widetilde{\varepsilon}_t + o_p(1)$ which shows that $T^{-1/2}\sum_{t=1}^{[sT]} x_t e_t$ is a functional of $T^{-1/2}\sum_{t=1}^{[sT]} x_t\widetilde{\varepsilon}_t$.

Step 2: Next I derive the limiting distribution of $T^{-1/2}\sum_{t=1}^{[sT]} x_t\widetilde{\varepsilon}_t$. Note that it can be written as

$$T^{-1/2}\sum_{t=1}^{[sT]} x_t\widetilde{\varepsilon}_t = T^{-1/2}\sum_{t=1}^{[sT]}\left[x_t\varepsilon_t + x_tx_t'(\beta_t-\beta_0) + x_tz_t'(\gamma_t-\gamma_0)\right]$$

The limit of the first term above directly follow from Condition 1(2), $T^{-1/2}\sum_{t=1}^{[sT]} x_t\varepsilon_t \Rightarrow \mathrm{Var}(x_t\varepsilon_t)^{1/2}W(s)$; The limit of the second follows from Conditions 1(1) and 1(3), $T^{-1/2}\sum_{t=1}^{[sT]} x_tx_t'(\beta_t-\beta_0) = T^{-1/2}\sum_{t=1}^{[sT]} x_tx_t'\lambda_\beta T^{-1}\sum_{i=1}^{t}\nu_{1i} \Rightarrow \lambda_\beta\Sigma_{xx}\Sigma_\beta^{1/2}\int_0^s W_\beta(r)dr$; and the limit of the third follows from Conditions 1(1) and 1(3), $T^{-1/2}\sum_{t=1}^{[sT]} x_tz_t'(\gamma_t-\gamma_0) = T^{-1/2}\sum_{t=1}^{[sT]} x_tz_t'\lambda_\gamma T^{-1}\sum_{i=1}^{t}\nu_{2i} \Rightarrow \lambda_\gamma\Sigma_{xz}\Sigma_\gamma^{1/2}\int_0^s W_\gamma(r)dr$. Therefore, limiting distributions of $T^{-1/2}\sum_{t=1}^{[sT]} x_t\widetilde{\varepsilon}_t$ and $T^{-1/2}\sum_{t=1}^{[sT]} x_te_t$ can be derived, and them are summerized in Table 7, together with the limiting distributions of other partial sums. Note that the absence of the λ_γ term in the limiting distribution of $T^{-1/2}\sum_{t=1}^{[sT]} u_te_t$ is a direct consequence of $\Sigma_{uz} = E(u_tz_t') = 0$ in the partial regression.

Finally, the asymptotics for the constant β model $y_t = x_t'\beta_0 + z_t'\gamma_t + \varepsilon_t$ can be obtained by setting $\lambda_\beta = 0$ in Table 7.

9.8 Proposition 4 (Section 4.2.2)

I will show the L_T^u version of the Nyblom test statistic and the Chow-based test statistics are functionals of $T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t$. Therefore their limiting distributions are functionals of the limit of $T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t$.

Step 1: For the L_T^u test, I will first show $T^{-1/2}\sum_{t=1}^{[sT]}\widehat{u}_te_t$ has the same limiting distribution as that of $T^{-1/2}\sum_{t=1}^{[sT]} u_te_t$. Define the following matrix notation for the subsamples, say, for $t\in(0,[sT])$, $Z_1 = (z_1,\cdots,z_{[sT]})'$, $U_1 = (u_1,\cdots,u_{[sT]})'$ and $e_1 = (e_1,\cdots,e_{[sT]})'$. Let $M_{z1} = I - Z_1(Z_1'Z_1)^{-1}Z_1'$ and $P_{z1} = I - M_{z1}$. Notation for the second subsample can be defined likewise. Then for the first subsample, $T^{-1/2}\sum_{t=1}^{[sT]}\widehat{u}_te_t = T^{-1/2}U_1'e_1 + T^{-1/2}U_1'P_{z1}e_1$. So the key is to show the second term vanishes in the limit. The second term can be written as $T^{-1/2}U_1'P_{z1}e_1 = T^{-1}U_1'Z_1\left[T^{-1}Z_1'Z_1\right]^{-1}T^{-1/2}Z_1'e_1$ where $T^{-1}U_1'Z_1 \xrightarrow{p} sE(u_tz_t') = 0$ which follows from Condition 1(1) and the orthogonal condition in partial regression. $\left[T^{-1}Z_1'Z_1\right]^{-1} \xrightarrow{p} s^{-1}[E(z_tz_t')]^{-1}$ following from Condition 1(1) and the continuous mapping theorem.. The distribution of $T^{-1/2}Z_1'e_1$ is given in Table 7. Therefore, $T^{-1/2}U_1'P_{z1}e_1 \to 0$, that is, the use of residual $\widehat{u}_t$ instead of u_t has not asymptotic effect, just as in the corresponding stable model.

$$T^{-1/2}\sum_{t=1}^{[sT]}\widehat{u}_te_t \;=\; T^{-1/2}\sum_{t=1}^{[sT]} u_te_t \;+\; o_p(1).$$

The connection between $T^{-1/2}\sum_{t=1}^{[sT]} u_te_t$ and $T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t$ is stated in Table 7.

Step 2: Since all Chow-based tests are functionals of the Chow test, it is sufficient to show the Chow test is a functional of $T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t$. Result (49) can be applied to subsamples which yields[75]

$$T^{1/2}\left(\widehat{\beta_1}-\beta_0\right) = \frac{1}{s}\Sigma_{uu}^{-1}\,T^{1/2}U_1'\widetilde{\varepsilon}_1 + o_p(1), \qquad T^{1/2}\left(\widehat{\beta_2}-\beta_0\right) = \frac{1}{1-s}\Sigma_{uu}^{-1}\,T^{1/2}U_2'\widetilde{\varepsilon}_2 + o_p(1)$$

Take the difference and rearrange,

$$T^{1/2}\left(\widehat{\beta_1}-\widehat{\beta_2}\right) \;=\; \frac{1}{s(1-s)}\Sigma_{uu}^{-1}\left[T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t - sT^{-1/2}\sum_{t=1}^{T} u_t\widetilde{\varepsilon}_t\right] + o_p(1)$$

[75] the appearance of Σ_{uu} in both subsample limits is the consequence of Condition 1(1). For example, for $t\in(0,[sT])$, $\frac{U_1'U_1}{sT} = s^{-1}T^{-1}\sum_{t=1}^{[sT]} u_tu_t' = s^{-1}s\Sigma_{uu} + o_p(1) \xrightarrow{p} \Sigma_{uu}$.

Therefore, the Chow statistic is a functional of $T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t$. So do all tests that are functionals of the Chow test statistic.

Step 3: The limiting distribution of $T^{-1/2}\sum_{t=1}^{[sT]} u_t\widetilde{\varepsilon}_t$ is given in Table 7. By the continuous mapping theorem, the distributions of all regression-based tests are easily obtained. This leads to Proposition 4.

9.9 Proposition 5 (Section 4.3)

Denote $W_T(s) = \text{Var}(u_t\varepsilon_t)^{-1/2}T^{-1/2}\sum_{t=1}^{T} u_t\widetilde{\varepsilon}_t$. To show the independence between the J-test and the stability tests, I proceed as follows.

1. J-test statistic can be expressed as a functional of $W_T(1)$.
2. The regression-based stability tests can be expressed as functionals of $W_T(s) - sW_T(1)$.
3. Under the null of $\{\beta_0, \gamma_t\}$, $W_T(1)$ and $W_T(s) - sW_T(1)$ are asymptotically independent.
4. Under the null hypothesis, the J-test and stability tests are asymptotically independent.

Step 1: $J_T = T^{-1}[W_0^{1/2}g(\widehat{\phi}, \theta_0)]'\bar{M}_\theta[W_0^{1/2}g(\widehat{\phi}, \theta_0)] + o_p(1)$ where $T^{-1/2}g(\widehat{\phi}, \theta_0) = T^{-1/2}D_g(\widehat{\phi} - \phi_0) + O_p(T^{-1/2}) = T^{-1/2}(\widehat{\beta} - \beta_0) + O_p(T^{-1/2})$ The second equality follows from (5). The last equality follows from the transfermation in Section 3.2. Then, by (49), $T^{1/2}(\widehat{\beta} - \beta_0)$ is a functional of $T^{-1/2}\sum_{t=1}^{T} u_t\widetilde{\varepsilon}_t$. So does J_T. Hence, J_T is a functional of $W_T(1)$, $J_T = J(W_T(1))$.

Step 2: Appendix 9.8 shows that the robust stability tests are functionals of $W_T(s) - sW_T(1)$. Let ξ_T denote any regression-based test. Then, $\xi_T = \xi(W_T(s) - sW_T(1))$.

Step 3: The limiting distribution of $W_T(s)$ is given in Table 7, according to which, the null distribution is $W_T(s) \Rightarrow W(s)$, where $W(s)$ is a k-dimensional standard Brownian motion with $k = \dim(\beta)$. Then under the null hypothesis,

$$W_T(1) \;\Rightarrow\; W(1), \qquad W_T(s) - sW_T(1) \;\Rightarrow\; W(s) - sW(1)$$

Notice that, $W(1) = W(s) + W(1-s)$ and $W(s) - sW(1) = (1-s)W(s) - sW(1-s)$ where $W(s) \sim N(0, sI)$ and $W(1-s) \sim N(0, (1-s)I)$. Given $W(s)$ and $W(1-s)$ are independent, $W(1)$ and $W(s) - sW(1)$ are independent because

$$\begin{aligned} & \text{Cov}(W(1),\ W(s) - sW(1)) \\ = \; & \text{Cov}(W(s) + W(1-s),\ (1-s)W(s) - sW(1-s)) \\ = \; & (1-s)\text{Var}(W(s)) - s\text{Var}(W(1-s)) \;=\; 0. \end{aligned}$$

Step 4: By the continuous mapping theorem, the joint distribution of $J(W_T(1))$ and $\xi(W_T(s) - sW_T(1))$ under the null hypothesis is

$$\begin{bmatrix} J(W_T(1)) \\ \xi(W_T(s) - sW_T(1)) \end{bmatrix} \Rightarrow \begin{bmatrix} J(W(1)) \\ \xi(W(s) - sW(1)) \end{bmatrix}$$

It follows from the asymptotic independence of $W_T(1)$ and $W_T(s) - sW_T(1)$ under the null and the continuous mapping theorem, the J-test statistic $J(W_T(1))$ and any regression-based stability test $\xi(W_T(s) - sW_T(1))$ are asymptotically independent under the null.

9.10 Proposition 6 (Section 5.2)

Under Condition 3, by a local linearization similar to (12) in the two-step model, the overall restriction (23) can be decomposed into its stable component and its TVP component. Let $A(\phi_0) = \partial a(\phi_0)/\partial\phi$.

$$\begin{aligned} \text{Restriction } 1: \quad 0 &= a(\phi_0) \\ \text{Restriction } 2: \quad 0 &= A(\phi_0)(\phi_t - \phi_0). \end{aligned}$$

9.10.1 Conventional tests for $a(\phi_0) = 0$

The likelihood ratio test, the Lagrange multiplier test and the Wald test can all be used to test for $a(\phi_0) = 0$. In what follows, I first derive the restricted and unrestriced estimators using the standard procedure (i.e., ignoring the time variation in ϕ_t when estimating ϕ). Denote the restriced and the unrestriced estimators by by $\widetilde{\phi}$ and $\widehat{\phi}$ respectively. Then I show under the null of $a(\phi_t) = 0$, the tests are independent of the time variation in ϕ_t.

The *restricted* estimator is obtained by minimizing the objective function Q_T s.t. $a(\phi) = 0$. Let γ_T be the Lagrange multiplier to the constrained problem. Then $\widetilde{\phi}$ satisfies the first order condition

$$\begin{aligned} T^{-1/2}\,\frac{\partial Q_T(\widetilde{\phi})}{\partial\phi} \;+\; T^{-1/2}\,A(\widetilde{\phi})'\gamma_T &= 0 \\ T^{-1/2}\,a(\widetilde{\phi}) &= 0 \end{aligned} \tag{55}$$

To solve for $\widetilde{\phi}$, the terms in the first order condition are linearized. Expanding $a(\widetilde{\phi})$ around the overall restriction $a(\phi_t)$ results in $a(\widetilde{\phi}) = a(\phi_t) + A(\phi_0)(\widetilde{\phi} - \phi_t) + O_p(T^{-1})$, where the order of the magnitude of the remainder is obtained under Condition 3 and $\widetilde{\phi} \xrightarrow{p} \phi_0$. Thus under the null of $a(\phi_t) = 0$,

$$T^{1/2}a(\widetilde{\phi}) \;=\; A(\phi_0)\,T^{1/2}(\widetilde{\phi} - \phi_0) - A(\phi_0)\,T^{1/2}(\phi_t - \phi_0) + O_p(T^{-1/2}) \tag{56}$$

It is straightforward to show $A(\widetilde{\phi}) = A(\phi_0) + O_p(T^{-1})$ under Condition 3. Therefore,

$$T^{1/2}A(\widetilde{\phi}) \;=\; T^{1/2}A(\phi_0) \;+\; O_p(T^{-1/2}). \tag{57}$$

Under Conditions 4(1) and 4(2), the Taylor expansion of $\partial Q_T(\widetilde{\phi})/\partial\phi$ around ϕ_t leads to

$$\begin{aligned} T^{1/2}\frac{\partial Q_T(\widetilde{\phi})}{\partial\phi} &= T^{-1/2}\sum s_t(\phi_t) \;+\; T^{-1/2}\sum h_t(\phi_0)(\widetilde{\phi} - \phi_t) \;+\; O_p(T^{-1/2}) \\ &= T^{-1/2}\sum s_t(\phi_t) \;+\; T^{-1/2}\sum h_t(\phi_0)(\widetilde{\phi} - \phi_0) \\ &\quad + T^{-1/2}\sum h_t(\phi_0)(\phi_t - \phi_0) \;+\; O_p(T^{-1/2}) \\ &= T^{-1/2}\sum s_t(\phi_t) \;+\; H_0\,T^{1/2}(\widetilde{\phi} - \phi_0) \;+\; H_0\,T^{-1/2}\sum(\phi_t - \phi_0) \;+\; o_p(1) \end{aligned} \tag{58}$$

where $H_0 = Eh_t(\phi_0)$. The last equality uses Conditions 4(1) and 4(3). To streamline the notation, let $A_0 = A(\phi_0)$. Substituting (56), (57) and (58) into the first order condition (55) and rearranging, yields

$$\begin{bmatrix} H_0 & A_0' \\ A_0 & 0 \end{bmatrix} \begin{bmatrix} T^{1/2}(\widetilde{\phi} - \phi_0) \\ T^{1/2}\gamma_T \end{bmatrix} = -\begin{bmatrix} T^{-1/2}\sum s_t(\phi_t) \\ 0 \end{bmatrix} + \begin{bmatrix} H_0\,T^{-1/2}\sum(\phi_t - \phi_0) \\ A_0\,T^{-1/2}\sum(\phi_t - \phi_0) \end{bmatrix} + o_p(1)$$

Calculate the inverse of the partitioned matrices, $\widetilde{\phi}$ and γ_T can be jointly solved as

$$\begin{bmatrix} T^{1/2}\ (\widetilde{\phi}-\phi_0) \\ T^{1/2}\ \gamma_T \end{bmatrix} = \begin{bmatrix} F_{11} & F_{12} \\ F_{21} & F_{22} \end{bmatrix} \left\{ - \begin{bmatrix} T^{-1/2}\sum s_t(\phi_t) \\ 0 \end{bmatrix} + \begin{bmatrix} H_0\ T^{-1/2}\sum(\phi_t-\phi_0) \\ A_0\ T^{-1/2}\sum(\phi_t-\phi_0) \end{bmatrix} \right\} + o_p(1)$$

where in the above partitioned matrix,

$$\begin{aligned} F_{11} &= H_0^{-1} - H_0^{-1}A_0'[A_0H_0^{-1}A_0']^{-1}A_0H_0^{-1} \\ F_{12} &= H_0^{-1}A_0'[A_0H_0^{-1}A_0']^{-1} \\ F_{21} &= [A_0H_0^{-1}A_0']^{-1}A_0H_0^{-1} \\ F_{22} &= -[A_0H_0^{-1}A_0']^{-1} \end{aligned}$$

Solving $\widetilde{\phi}$ seperately yields

$$\begin{aligned} T^{1/2}(\widetilde{\phi}-\phi_0) &= -F_{11}T^{-1/2}\sum s_t(\phi_t) + (F_{11}H_0+F_{12}A_0)T^{-1/2}\sum(\phi_t-\phi_0) + o_p(1) \\ &= -[H_0^{-1} - H_0^{-1}A_0'(A_0H_0^{-1}A_0')^{-1}A_0H_0^{-1}]T^{-1/2}\sum s_t(\phi_t) \\ &\quad +T^{-1/2}\sum(\phi_t-\phi_0) + o_p(1) \end{aligned} \tag{59}$$

which follows from $F_{11}H_0 + F12A_0 = I$. Note that the second right-hand-side term in (59) represents the effect of the ignored instability in ϕ_t. Similarly, γ_T can be solved as

$$\begin{aligned} T^{1/2}\gamma_T &= -F_{21}T^{-1/2}\sum s_t(\phi_t) + (F_{21}H_0+F_{22}A_0)T^{-1/2}\sum(\phi_t-\phi_0) + o_p(1) \\ &= -[(A_0H_0^{-1}A_0']^{-1}A_0H_0^{-1}T^{-1/2}\sum s_t(\phi_t) + o_p(1) \end{aligned} \tag{60}$$

which follows from $F_{21}H_0 + F22A_0 = 0$. Note that the instability in ϕ_t has no effects on γ_T.

The derivation of the unrestricted estimator $\widehat{\phi}$ is similar to that in the linear regression model. Hence it is not repeated here. $\widehat{\phi}$ can be solved as

$$T^{1/2}(\widehat{\phi}-\phi_0) \ = \ -H_0^{-1}T^{-1/2}\sum s_t(\phi_t) + T^{-1/2}\sum(\phi_t-\phi_0) + o_p(1) \tag{61}$$

Next, using the results for $\widetilde{\phi}$, $\widehat{\phi}$ and γ_T, I study the behavior of three tests. Denote the likelihood ratio test, the Lagrange multiplier test and the Wald test by LR_T, LM_T and W_T respectively.

(a) For the *likelihood ratio* test,

$$LR_T \ = \ 2T(Q_T(\widetilde{\phi}) - Q_T(\widehat{\phi})) \ = \ T(\widetilde{\phi}-\widehat{\phi})'H_0(\widetilde{\phi}-\widehat{\phi}) + o_p(T^{-1})$$

and its asymptotics depend on the behavior of $\widetilde{\phi}-\widehat{\phi}$. By (59) and (61),

$$\begin{aligned} T^{1/2}(\widehat{\phi}-\widetilde{\phi}) &= -H_0^{-1}A_0'[A_0H_0^{-1}A_0']^{-1}A_0H_0^{-1}T^{-1/2}\sum s_t(\phi_t) \\ &\quad +\ T^{-1/2}\sum(\phi_t-\phi_0) - T^{-1/2}\sum(\phi_t-\phi_0) + o_p(1) \\ &= -H_0^{-1}A_0'[A_0H_0^{-1}A_0']^{-1}A_0H_0^{-1}T^{-1/2}\sum s_t(\phi_t)\ +\ o_p(1) \end{aligned}$$

So $\widetilde{\phi}-\widehat{\phi}$ is a functional of $T^{-1/2}\sum s_t(\phi_t)$ only. There are no asymptotic effects caused by the instability in γ_t because the distortion on $\widetilde{\phi}$ and that on $\widehat{\phi}$ are of the same magnitude but with opposite signs.

(b) For the *Lagrange multiplier* test, by definition, $LM_T = T\ \gamma_T'\ \widehat{\text{Avar}}(\gamma_T)^{-1}\ \gamma_T$ where $\text{Avar}(\gamma_T)$ is the asymptotic variance of γ_T. From (60), the effects of the instability in ϕ_t on γ_T is canceled out and the distribution of γ_T only depends on $T^{-1/2}\sum s_t(\phi_t)$. So does LM_T.

(c) The *Wald statistic* is defined as $W_T = T\ a(\widehat{\phi})'\ \widehat{\text{Avar}}(a(\widehat{\phi}))^{-1}\ a(\widehat{\phi})$ where $\text{Avar}(a(\widehat{\phi}))$ is the asymptotic variance of $a(\widehat{\phi})$. Taylor expansion of $a(\widehat{\phi})$ around $a(\phi_t)$ yields $a(\widehat{\phi}) = a(\phi_t) + A_0(\widehat{\phi} - \phi_t) + O_p(T^{-1/2})$ which follows Condition 3 and the consistency of $\widehat{\phi}$. Since $a(\phi_t) = 0$ under the null hypothesis,

$$T^{1/2}a(\widehat{\phi}) = A_0 T^{1/2}(\widehat{\phi} - \phi_0) - A_0 T^{-1/2}\sum(\phi_t - \phi_0) + O_p(T^{-1/2}) \tag{62}$$

Substituting (61) into (62) yields

$$\begin{aligned} T^{1/2}a(\widehat{\phi}) &= A_0\left[-H_0^{-1}T^{-1/2}\sum s_t(\phi_t) + T^{-1/2}\sum(\phi_t - \phi_0)\right] \\ &\quad -A_0 T^{-1/2}\sum(\phi_t - \phi_0) + O_p(T^{-1/2}) \\ &= -A_0 H_0^{-1}T^{-1/2}\sum s_t(\phi_t) + o_p(1) \end{aligned}$$

Thus asymptotically $a(\widehat{\phi})$, hence W_T, is unaffected by ϕ_t.

In summary, the distribution of either the unrestriced or the restricted estimator of ϕ_0 is affected by the instability in ϕ_t. However, no matter what statistic is used to test for $a(\phi_0) = 0$, it is a functional of $T^{-1/2}\sum s_t(\phi_t)$ only. The distortion term $T^{-1/2}\sum(\phi_t - \phi_0)$ is always canceled out in these test statistics.

9.10.2 Size control of the joint test

Let $W_T(s) = \Omega^{-1/2}T^{-1/2}\sum_{t=1}^{[sT]} s_t(\phi_t)$ where Ω is the asymptotic variance of $s_t(\phi_t)$ defined in Condition 4(2). To show the independence between the conventional specification tests for $a(\phi_0) = 0$ and the stabillity tests for $A(\phi_0)(\phi_t - \phi_0) = 0$, I derive in four steps.

1. Tests for $a(\phi_0) = 0$ are functionals of $W_T(1)$ under the null.
2. The Chow-based stability tests are functionals of $W_T(s) - sW_T(1)$ under the null.
3. $W_T(1)$ and $W_T(s) - sW_T(1)$ are asymptotically independent.
4. The tests for $a(\phi_0) = 0$ and stability tests are asymptotically independent.

The first step is shown in Part I . Now turn to the second step. The class of tests considered are based on sequntial Chow tests. The Chow test is built up on the estimated difference of $A(\phi)\phi$ over different sample periods, that is, based on $A(\widehat{\phi}_1)\widehat{\phi}_1 - A(\widehat{\phi}_2)\widehat{\phi}_2$. Applying (61) to the subsamples for any $0 \leq s \leq 1$, yields

$$T^{1/2}(\widehat{\phi}_1 - \phi_0) = \frac{1}{s}\left[-H_0^{-1}\ T^{-1/2}\sum_{t=1}^{[sT]} s_t(\phi_t)\ +\ T^{-1/2}\sum_{t=1}^{[sT]}(\phi_t - \phi_0)\right] + o_p(1) \tag{63}$$

$$T^{1/2}(\widehat{\phi}_2 - \phi_0) = \frac{1}{1-s}\left[-H_0^{-1}\ T^{-1/2}\sum_{t=[sT]+1}^{T} s_t(\phi_t)\ +\ T^{-1/2}\sum_{t=[sT]+1}^{T}(\phi_t - \phi_0)\right] + o_p(1)$$

Note that because of the consistency of $\widehat{\phi}_i$, for $i = 1$ and 2, and the continuity of function $A(\phi)$ assumed in Condition 3, I obtain $T^{1/2}A(\widehat{\phi}_i)(\widehat{\phi}_i - \phi_0) \; = \; T^{1/2}A(\phi_0)(\widehat{\phi}_i - \phi_0) + o_p(1)$. Then substitute into (63),

$$
\begin{aligned}
& T^{1/2}\;(A(\widehat{\phi}_1)\widehat{\phi}_1 - A(\widehat{\phi}_2)\widehat{\phi}_2) \; = \; T^{1/2}(A_0\widehat{\phi}_1 - A_0\widehat{\phi}_2) + o_p(1) \\
= \quad & \frac{-1}{s(1-s)} A_0 H_0^{-1} \left[T^{-1/2} \sum_{t=1}^{[sT]} s_t(\phi_t) - sT^{-1/2} \sum_{1}^{T} s_t(\phi_t) \right] \\
& + \frac{-1}{s(1-s)} \left[T^{-1/2} \sum_{t=1}^{[sT]} A_0(\phi_t - \phi_0) - sT^{-1/2} \sum_{1}^{T} A_0(\phi_t - \phi_0) \right] + o_p(1) \\
= \quad & \frac{-1}{s(1-s)} A_0 H_0^{-1} \left[T^{-1/2} \sum_{t=1}^{[sT]} s_t(\phi_t) - sT^{-1/2} \sum_{1}^{T} s_t(\phi_t) \right] + o_p(1)
\end{aligned}
$$

where I use that $A_0(\phi_t - \phi_0) = 0$ for all t under the null. It is evident $A(\widehat{\phi}_1)\widehat{\phi}_1 - A(\widehat{\phi}_2)\widehat{\phi}_2$ is a functional of $W_T(s) - sW_T(1)$. Hence any stability tests based on Chow test are functionals of $W_T(s) - sW_T(1)$.

Finally note that under Condition 3(2), $W_T(s) \Rightarrow W(s)$ where W(s) is a standard Brownian motion. The derivation in Appendix 7.10 for Step 3 and Step 4 then carries over.

9.11 Derivation of the Investment Equations (Section 6.2.3)

I derive in details the investment model in ΔI. Derivation of the model in IK and the model in ΔK follows the same procedure, and hence is not repeated.

Consider the cost function (30). The partial derivative of the period profit function defined in (26) with respect to I_t, I_{t-1} and K_t, needed to specify the investment Euler equation, are

$$\frac{\partial R_t}{\partial I_t} = -\left[\frac{\partial C(I_t, I_{t-1})}{\partial I_t} + 1\right] p_t = -\left[1 + \phi_0 + 2\phi_1 \frac{I_t}{I_{t-1}} + 3\phi_2 \frac{I_t^2}{I_{t-1}^2}\right] p_t \tag{64}$$

$$\frac{\partial R_t}{\partial I_{t-1}} = -\frac{\partial C(I_t, I_{t-1})}{\partial I_{t-1}} p_t = \left[\phi_1 \frac{I_t^2}{I_{t-1}^2} + 2\phi_2 \frac{I_t^3}{I_{t-1}^3}\right] p_t \tag{65}$$

$$\frac{\partial R_t}{\partial K_t} = \frac{\partial F(K_t, L_t)}{\partial K_t} = \alpha \frac{Y_t}{K_t} \tag{66}$$

The firm's optimization problem is then to choose processes I_t, K_t and L_t for all dates $t \geq 0$ to maximize (25) subject to (24), (28), (26) and (27). To carry out this constrained optimization, define the Lagrangian

$$\mathcal{L}_t = E_t \left\{ \sum_{s=t}^{\infty} \beta^{s-t} [R_s - \lambda_s (K_{s+1} - (1-\delta) K_s - I_s)] \right\}$$

where λ_s is the langragian multiplier for period s. Setting $\partial \mathcal{L}_t / \partial x_s = 0$ with $x_s = [I_s \; K_s \; L_s]'$, yields the first-order conditions. In particular, for $s = t$, the first-order conditions for I_t and K_t are

$$[I]: \qquad 0 = \lambda_t + \frac{\partial R_t}{\partial I_t} + \beta E_t \left[\frac{\partial R_{t+1}}{\partial I_t}\right] \tag{67}$$

$$[K]: \qquad 0 = \lambda_t - \beta(1-\delta) E_t \lambda_{t+1} - \beta E_t \left[\frac{\partial R_{t+1}}{\partial K_{t+1}}\right] \tag{68}$$

Combining (67) and (68) to eliminate λ_t and λ_{t+1}, yields

$$\frac{\partial R_t}{\partial I_t} + \beta E_t \left[\frac{\partial R_{t+1}}{\partial I_t}\right] = \beta(1-\delta) E_t \left[\frac{\partial R_{t+1}}{\partial I_{t+1}}\right] + \beta^2 (1-\delta) E_t \left[\frac{\partial R_{t+2}}{\partial I_{t+1}}\right] - \beta E_t \left[\frac{\partial R_{t+1}}{\partial K_{t+1}}\right] \tag{69}$$

Substituting (64), (65) and (66) into (69), the Euler equation of investment is

$$\begin{aligned}
& \left[1 + \phi_0 + 2\phi_1 \frac{I_t}{I_{t-1}} + 3\phi_2 \frac{I_t^2}{I_{t-1}^2}\right] p_t \\
= \;& \beta E_t \left\{ \left[(1-\delta)(1+\phi_0) + 2(1-\delta)\phi_1 \frac{I_{t+1}}{I_t} + (\phi_1 + 3(1-\delta)\phi_2) \frac{I_{t+1}^2}{I_t^2} + 2\phi_2 \frac{I_{t+1}^3}{I_t^3}\right] p_{t+1} \right\} \\
& -\beta^2 (1-\delta) E_t \left\{ \left[\phi_1 \frac{I_{t+2}^2}{I_{t+1}^2} + 2\phi_2 \frac{I_{t+2}^3}{I_{t+1}^3}\right] p_{t+2} \right\} + \alpha \beta E_t \left[\frac{Y_{t+1}}{K_{t+1}}\right]
\end{aligned}$$

Log-linearizing the above Euler equation around the stationary steady-state equilibrium gives

$$\begin{aligned}
\Delta \widehat{I}_t = \;& \beta(2-\delta) E_t \Delta \widehat{I}_{t+1} - \beta^2 (1-\delta) E_t \Delta \widehat{I}_{t+2} \\
& + \gamma_1 \widehat{p}_t + \gamma_2 E_t \widehat{p}_{t+1} + \gamma_3 E_t \widehat{p}_{t+2} + \gamma_4 E_t \widehat{KY}_{t+1}
\end{aligned} \tag{70}$$

where

$$
\begin{aligned}
\gamma_1 &= -\frac{1+\phi_0+2\phi_1+3\phi_2}{2(\phi_1+3\phi_2)} \\
\gamma_2 &= \frac{\beta[(1-\delta)(1+\phi_0)+(3-2\delta)\phi_1+(5-3\delta)\phi_2]}{2(\phi_1+3\phi_2)} \\
\gamma_3 &= -\frac{\beta^2(1-\delta)(\phi_1+2\phi_2)}{2(\phi_1+3\phi_2)} \\
\gamma_4 &= -\frac{[1-\beta(1-\delta)](1+\phi_0)+[2-\beta(3-2\delta)+\beta^2(1-\delta)]\phi_1+[3-\beta(5-3\delta)+2\beta^2(1-\delta)]\phi_2}{2(\phi_1+3\phi_2)}
\end{aligned}
$$

By the expressions of γ_1, γ_2 and γ_3, their sum is very close to zero[76]. Then substitute $\gamma_2 \approx -(\gamma_2+\gamma_3)$ into equation (70), I get

$$
\Delta\widehat{I}_t \approx \beta(2-\delta)E_t\Delta\widehat{I}_{t+1} - \beta^2(1-\delta)E_t\Delta\widehat{I}_{t+2} - \gamma_1 E_t\Delta\widehat{p}_{t+1} + \gamma_3 E_t\Delta\widehat{p}_{t+2} + \gamma_4 E_t\widehat{KY}_{t+1} \tag{71}
$$

Following the identical procedure, when the cost adjustment function is (31), the log-linearized investment Euler equation is

$$
\begin{aligned}
\Delta\widehat{I}_t &= \beta(2-\delta)E_t\Delta\widehat{I}_{t+1} - \beta^2(1-\delta)E_t\Delta\widehat{I}_{t+2} \qquad (72)\\
&\quad +\lambda_1[\beta(1-\delta)E_t\widehat{I}_{t+1} - \widehat{I}_t] + \lambda_2[\beta(1-\delta)E_t\widehat{p}_{t+1} - \widehat{p}_t] + \lambda_3 E_t\widehat{KY}_{t+1}
\end{aligned}
$$

with $\lambda_1 = \widetilde{\phi}_2/\widetilde{\phi}_4$, $\lambda_2 = \bar{p}/(\widetilde{\phi}_4\bar{I})$ and $\lambda_3 = \alpha\beta\bar{Y}/(\widetilde{\phi}_4\bar{K}\bar{I})$ where a bar over a variable is the steady-state value of that variable. Since $\beta(1-\delta)$ is very close to one for theoretical values of β and δ, equation (72) can be approximated by

$$
\Delta\widehat{I}_t \approx [\beta(2-\delta)+\lambda_1]E_t\Delta\widehat{I}_{t+1} - \beta^2(1-\delta)E_t\Delta\widehat{I}_{t+2} + \lambda_2 E_t\Delta\widehat{p}_{t+1} + \lambda_3 E_t\widehat{KY}_{t+1} \tag{73}
$$

Inspection of equations (71) and (73), they can be nested within the following specification

$$
\Delta\widehat{I}_t = \theta_{I1}E_t\Delta\widehat{I}_{t+1} - \beta^2(1-\delta)E_t\Delta\widehat{I}_{t+2} + \theta_{p1}E_t\Delta\widehat{p}^I_{t+1} + \theta_{p2}E_t\Delta\widehat{p}^I_{t+2} + \theta_{KY}E_t\widehat{KY}_{t+1} + \varepsilon_t
$$

which is the third equation in (36), the model in ΔI.

In addition, from the above derivation, the signs and, in some cases, the values of the parameters of the investment equations can be predicted. See Table 8 for a summary.

9.12 Cross-Equation Restrictions (Section 6.3)

I derive the cross-equation restrictions for model in ΔI. Derivation of the cross-equation restrictions for the model in IK and the model in ΔK follows the same procedure, and hence is not repeated.

The first-step VAR takes the form of $Z_t = \Phi Z_{t-1} + u_t$. Let $\Delta\widehat{I}_t$, $\Delta\widehat{p}_t$ and $\widehat{KY}_t$ take the first, the second and the third positions in Z_t respectively. The k-period ahead forecasts of the three variables are

$$
E(\Delta\widehat{I}_{t+k}|Z_t) = e_I'\Phi^k Z_t, \quad E(\Delta\widehat{p}_{t+k}|Z_t) = e_p'\Phi^k Z_t \quad \text{and} \quad E(\widehat{KY}_{t+k}|Z_t) = e_{KY}'\Phi^k Z_t \tag{74}
$$

where $e_I = [1\ 0\ 0\cdots 0]'$, $e_p = [0\ 1\ 0\cdots 0]'$ and $e_{ky} = [0\ 0\ 1\cdots 0]'$. Assuming $E(\varepsilon_t|Z_{t-1}) = 0$ where ε_t is the disturbance term in the investment Euler equation.

[76] To see this, when β is one and δ is zero, the sum of γ_1, γ_2 and γ_3 is identically zero. For theoretical values of β and δ used in the paper, that is, $\beta = 0.99$ and $\delta = 0.025$, we get $\gamma_1+\gamma_2+\gamma_3 = -0.017(1+\phi_0+\phi_1+\phi_2)/(\phi_1+2\phi_2)$.

Table 8: Theoretical Signs and Values of Coefficients

Model in ΔI		Model in IK		Model in ΔK	
Coeff.	Value/Sign	Coeff.	Value/Sign	Coeff.	Value/Sign
θ_{I1}	$+$	θ_{IK}	$+$ 0.99	θ_K	$+$ 0.99
θ_{I2}	$-$ 0.96	—	—	—	—
θ_{p1}	$+$	θ_{p1}	$+$	θ_{p1}	$+$
θ_{KY}	$+$	θ_{KY}	$+$	θ_{KY}	$+$

Note: When the value of a coefficient is not available, only the sign of that coefficient is reported. The values reported are computed by assuming $\beta = 0.99$ and $\delta = 0.025$.

Consider the third equation in (36), the model in ΔI. Taking expectations conditional on Z_{t-1} on both sides yields

$$\begin{aligned} E(\Delta\widehat{I}_t|Z_{t-1}) &= \theta_{I1}E_t(\Delta\widehat{I}_{t+1}|Z_{t-1}) + \theta_{I2}E_t(\Delta\widehat{I}_{t+2}|Z_{t-1}) \\ &\quad +\theta_{p1}E_t(\Delta\widehat{p}^I_{t+1}|Z_{t-1}) + \theta_{p2}E_t(\Delta\widehat{p}^I_{t+2}|Z_{t-1}) + \theta_{KY}E_t(\widehat{KY}_{t+1}|Z_{t-1}) \end{aligned} \tag{75}$$

Note expression (75) holds exactly, as a result of the orthogonal condition $E(\varepsilon_t|Z_{t-1}) = 0$ and the law of iterative expectations. Then using VAR forecasts (74), equation (75) becomes

$$(e'_I\Phi - \theta_{I1}e'_I\Phi^2 - \theta_{I2}e'_I\Phi^3 - \theta_{p1}e'_p\Phi^2 - \theta_{p2}e'_p\Phi^3 - \theta_{KY}e'_{KY}\Phi^2)Z_{t-1} = 0.$$

Since the above equality holds for any t, it must be the case that

$$e'_I\Phi - \theta_{I1}e'_I\Phi^2 - \theta_{I2}e'_I\Phi^3 - \theta_{p1}e'_p\Phi^2 - \theta_{p2}e'_p\Phi^3 - \theta_{KY}e'_{KY}\Phi^2 = 0$$

which is a $1 \times np$ vector of restrictions. So as long Φ^{-1} exists, which is indeed the case for the VAR specification in my application, one can post-multiply the above equation by Φ^{-1}. Then expressing the resulting restrictions in a column vector gives

$$\begin{aligned} 0 &= e_I - \Phi' e_I\theta_{I1} - (\Phi')^2 e_I\theta_{I2} - \Phi' e_p\theta_{p1} - (\Phi')^2 e_p\theta_{p2} - \Phi' e_{KY}\theta_{KY} \\ &= \left[I - \Phi'\theta_{I1} - (\Phi')^2\theta_{I2}\right] e_I - \left[\Phi'\theta_{p1} - (\Phi')^2\theta_{p2}\right] e_p - \Phi'\theta_{KY}\ e_{KY} \end{aligned}$$

9.13 Procedure for Estimation and Testing (Section 6.3)

The cross-equation restricions for the models investigated in the paper are linear in the second-step parameters, see (37). Hence, they can be written in the following general form

$$0 \;=\; A(\Phi_t) + B(\Phi_t)\theta. \tag{76}$$

Expressions of $A(\Phi_t)$, $B(\Phi_t)$ and θ for various models are listed in Table 9.

Table 9: Expressions of $A(\Phi_t)$, $B(\Phi_t)$ and θ

Model	$A(\Phi_t)$	$B(\Phi_t)$	θ
Model in IK	e_{IK}	$-[\Phi_t' e_{IK}\ \Phi_t' e_p\ \Phi_t' e_{KY}]$	$[\theta_{IK}\ \theta_{p1}\ \theta_{KY}]'$
Model in ΔK	e_K	$-[\Phi_t' e_K\ \Phi_t' e_p\ \Phi_t' e_{KY}]$	$[\theta_K\ \theta_{p1}\ \theta_{KY}]'$
Model in ΔI	e_I	$-[\Phi_t' e_I\ (\Phi_t')^2 e_I\ \Phi_t' e_p\ (\Phi_t')^2 e_p\ \Phi_t' e_{KY}]$	$[\theta_{I1}\ \theta_{I2}\ \theta_{p1}\ \theta_{p2}\ \theta_{KY}]'$

Then, following the analysis in Section 3.1, the overall restriction (76) can be decomposed into its time-invariant component and the TVP component results in

$$\text{Restriction 1}: \quad 0 = A(\Phi_0) + B(\Phi_0)\,\theta_0 \tag{77}$$

$$\text{Restriction 2}: \quad 0 = D_g\ \text{vec}(\Phi_t - \Phi_0)' \tag{78}$$

where expressions of D_g for the model in IK, the model in ΔK and the model in ΔI are respectively

$$\begin{aligned} D_g IK &= -[e_{IK}'\theta_{IK} + +e_p'\theta_{p1} + e_{KY}'\theta_{KY}]' \otimes I \\ D_g \Delta K &= -[e_K'\theta_K + +e_p'\theta_{p1} + e_{KY}'\theta_{KY}]' \otimes I \\ D_g \Delta I &= -[e_I'(\theta_{I1} + \theta_{I2}\Phi_0) + e_p'(\theta_{p1} + \theta_{p2}\Phi_0) + e_{KY}'\theta_{KY}]' \otimes I - (e_I'\theta_{I2} + e_p'\theta_{p2}) \otimes \Phi_0' \end{aligned}$$

Then, estimation and testing can be conducted as follows.

- *Step 1:* Obtain $\widehat{\Phi}$, the OLS estimator of Φ_0 through equation-by-equation estimation. Compute the residuals $\{e_t\}$.

- *Step 2:* Use $\widehat{\Phi}$ and solve for θ by the conventional two-step minimum distance approach. Obtain the efficient estimator $\widehat{\theta}$ and its estimated asymptotic variance $\widehat{\text{Avar}}(\widehat{\theta})$.

- *Step 3:* Compute the J-statistic J_T for testing the over-identifying restriction $A(\Phi_0)+B(\Phi_0)\theta_0 = 0$ and compute the p-value $P_{R1} = P_{\chi^2}(J > J_T)$.

- *Step 4:* To conduct the stability test, rewrite the VAR in the form of (**??**) by setting $\widehat{\beta} = D_g(\widehat{\Phi}, \widehat{\theta})\ \text{vec}\widehat{\Phi}'$. Obtain $\{x_t\}$ and $\{z_t\}$ accordingly.

- *Step 5:* Compute the residuals $\{\widehat{u}_t\}$ by regressing x_t on z_t. Together with $\{e_t\}$, construct $\widehat{\Sigma}_{uu} = T^{-1}\sum \widehat{u}_t\widehat{u}_t'$ and $\text{Var}\widehat{(u_t\varepsilon_t)} = \widehat{\Sigma}_{uu}(T^{-1}e_t^2)$.

- *Step 6:* Compute the test statistics for Restriction 2. The Nyblom statistic L_T^u can be computed according to (23). To obtain the Chow F_T-based tests, compute the F_T statistic by (22) and QLR_T, MW_T and EW_T by (21)[77].

[77]One needs to choose the trimming proportions s_0 and s_1, with $s_0 < s_1$. For each potential break date $t \in [s_0T,\ s_1T]$, calculate $\widehat{\beta}_i$ $(i = 1, 2)$ for subsample i by $\widehat{\beta}_i = D_g(\widehat{\Phi}_i, \widehat{\theta}_i)\ \text{vec}\widehat{\Phi}_i'$, or alternatively, by $\widehat{\beta}_i = D_g(\widehat{\Phi}, \widehat{\theta})\ \text{vec}\widehat{\Phi}_i'$ using the full-sample estimate of D_g.

- *Step 7:* Simulate the null distribution of the tests[78]. Compute the p-value $P_{R2} = P_{\lambda_1=0}(\xi > \xi_T)$, where ξ_T can be either of the L_T^u, QLR_T, MW_T or EW_T statistics obtained in Step 6.

- *Step 8:* For a given significance level of the overall test for (76), determine the significance levels of the R1 test and the R2 test, α_{R1} and α_{R2}. The decision rule is: the NKPC is not rejected if and only if $P_{R_1} > \alpha_{R_1}$ and $P_{R_2} > \alpha_{R_2}$.

9.14 Data Appendix

Table 10: Data Description

DRI name	Description
FYFF	Interest rate: federal funds (effective)
GDPD	Implicit price deflator: GDP
GDIPD	Implicit price deflator: producers durable equipment
LBGDPU	Implicit price deflator: non farm business
GDPQ	Gross domestic product: chained
GPBUQ	Gross domestic product: non farm business
GIPDEQ	Total producers durable equipment
KNNREQ	Real net stock, non-residential equipment

9.14.1 Data description

All data series in Table 10 are from DRI-McGraw Hill database.

- *Investment*, denoted by I in the paper, is the quarterly spending on producers durable equipment in billions of 1987 dollars.

- *Capital stock*, denoted by K in the paper, is a series interpolated from the annual net stock of private non-residential equipment. To construct this series, I set the fourth-quarter value of the interpolated series equal to the year-end value of net stock of that year. Then to interpolate between the year-end values, I assume the capital stock evolves following the law of motion $K_{t+1} = (1-\delta)K_t + I_t$ where δ is set to equal 0.025 on a quarterly basis. In words, the quarterly changes in the stock of equipment after depreciation is proportional to the quarterly pattern of investment. The resulting quarterly capital series is measured in billions of 1987 dollars.

- *Investment share in captial stock*, denoted by IK in the paper, is constructed as the ratio of investment to capital stock, I/K.

[78]For the QLR test, the MW test and the EW test, the critical values are available in Andrews (1993) and Andrews and Ploberger (1994). So alternatively one can compare the test statistics obtained in Step 6 with the critical values in these papers.

- *Purchase price of capital goods*, denoted by p^I in the paper, is constructed as $p^I = PDE/PY$, where PDE is the implicit price deflator for producers durable equipment and PY is the implicit price deflator of output. Two measures of PY are used. They are the price deflator for the non-farm business sector and the price deflator for the overall GDP measure. Accordingly, there are two measures for the relative purchase price of capital goods, denoted by p^I_{NFB} and p^I_{GDP} respectively.

- *Output*, is denoted by Y in the paper. Two measures of Y are used. One is the gross domestic product in the non-farm business sector, and the other one is an overall measure of gross domestic product. The two measures are denoted by Y_{NFB} and Y_{GDP} respectively.

- *Capital share in output*, denoted by KY in the paper, is computed as the ratio of capital stock to output, K/Y. Two measures of capital share are computed and are denoted by KY_{NFB} and KY_{GDP} respectively, corresponding to the two measures of output.

- *Short-term interest rate*, denoted by i in the paper, is the federal funds rate. Since only monthly series is available, it is converted to quarterly basis.

9.14.2 Persistence and data transformations

Almost all variables introduced above are used in their log-linearized form. The only exception is the interest rate, which is used without log-linearization. a log-linearized variable is computed as $\widehat{x}_t = \log x_t - T^{-1}\sum_{t=1}^{T}\log x_t$, that is, the log-deviation of a variable from its sample mean.

Table 11: Persistence of Series

Series	OLS	MUE	90% CI
i	0.92	0.93	0.87 $-$ 1.02
$\widehat{I}$	0.91	0.94	0.83 $-$ 1.02
$\widehat{K}$	0.96	0.99	0.88 $-$ 1.02
$\widehat{IK}$	0.95	0.95	0.87 $-$ 1.02
$\widehat{p}^I_{NFB}$	0.99	1.02	0.96 $-$ 1.03
$\widehat{p}^I_{GDP}$	0.99	1.02	0.95 $-$ 1.03
$\widehat{Y}_{NFB}$	0.88	0.89	0.80 $-$ 1.01
$\widehat{Y}_{GDP}$	0.88	0.89	0.79 $-$ 1.01
$\widehat{KY}_{NFB}$	0.96	0.97	0.89 $-$ 1.02
$\widehat{KY}_{GDP}$	0.96	0.99	0.91 $-$ 1.03

In addition, special attention is paid to the persistence of the time series, since persistence of series could lead to inference problem. It turns out that in my application, all variables in level, $\widehat{I}$, $\widehat{p^I}$, $\widehat{K}$, $\widehat{IK}$, $\widehat{KY}$, $\widehat{Y}$ and i, are persistent. Empirical estimates of persistence, measured by the largest autoregressive root of the series are give in Table 11. A second median-unbiased estimator (MUE) for the largest root is also reported[79]. They are construced by inverting the Dicker-Fuller unit root statistics (including a constant for the federal funds rate series, and including a linear time trend for the other series). Indeed, the hypothesis of a unit root cannot be rejected for any of the series.

To avoid the inference problem caused by persistent series, transformations are made to the time series. *First difference* is used for a series if (1) this variable shows up in an investment equation in first difference, such as investment $\widehat{I}$, capital stock $\widehat{K}$ and price of capital stock $\widehat{p}^I$; or (2) this variable only shows up in the first-step VAR model but not in any investment Euler equations, such as interest rate i and total output $\widehat{Y}$.

On the other hand, if a variable is persistent but showing up in an investment equation in level, such as capital share in output $\widehat{KY}$ and investment share in capital $\widehat{IK}$, then this variable is detrended using a non-parametric method, *the cubic spline*[80]. In the current application, a cubic spline with two knot points is used. Between the knot points, the spline is a third degree polynomial imposing equality of the levels and first two derivatives at the knot points. The knot points used are equally spaced.

Table 12: Variable Combinations of VARs

Dataset Name	VAR with interest rate	VAR without interest rate
$\Delta I(NFB)$	$[\Delta\widehat{I}\ \ \Delta\widehat{p}^I_{NFB}\ \ \widehat{KY}_{NFB}\ \ \Delta\widehat{Y}_{NFB}\ \ \Delta i]$	$[\Delta\widehat{I}\ \ \Delta\widehat{p}^I_{NFB}\ \ \widehat{KY}_{NFB}\ \ \Delta\widehat{Y}_{NFB}]$
$\Delta I(GDP)$	$[\Delta\widehat{I}\ \ \Delta\widehat{p}^I_{GDP}\ \ \widehat{KY}_{GDP}\ \ \Delta\widehat{Y}_{GDP}\ \ \Delta i]$	$[\Delta\widehat{I}\ \ \Delta\widehat{p}^I_{GDP}\ \ \widehat{KY}_{GDP}\ \ \Delta\widehat{Y}_{GDP}]$
$IK(NFB)$	$[\widehat{IK}\ \ \Delta\widehat{p}^I_{NFB}\ \ \widehat{KY}_{NFB}\ \ \Delta\widehat{Y}_{NFB}\ \ \Delta i]$	$[\widehat{IK}\ \ \Delta\widehat{p}^I_{NFB}\ \ \widehat{KY}_{NFB}\ \ \Delta\widehat{Y}_{NFB}]$
$IK(GDP)$	$[\widehat{IK}\ \ \Delta\widehat{p}^I_{GDP}\ \ \widehat{KY}_{GDP}\ \ \Delta\widehat{Y}_{GDP}\ \ \Delta i]$	$[\widehat{IK}\ \ \Delta\widehat{p}^I_{GDP}\ \ \widehat{KY}_{GDP}\ \ \Delta\widehat{Y}_{GDP}]$
$\Delta K(NFB)$	$[\Delta\widehat{K}\ \ \Delta\widehat{p}^I_{NFB}\ \ \widehat{KY}_{NFB}\ \ \Delta\widehat{Y}_{NFB}\ \ \Delta i]$	$[\Delta\widehat{K}\ \ \Delta\widehat{p}^I_{NFB}\ \ \widehat{KY}_{NFB}\ \ \Delta\widehat{Y}_{NFB}]$
$\Delta K(GDP)$	$[\Delta\widehat{K}\ \ \Delta\widehat{p}^I_{GDP}\ \ \widehat{KY}_{GDP}\ \ \Delta\widehat{Y}_{GDP}\ \ \Delta i]$	$[\Delta\widehat{K}\ \ \Delta\widehat{p}^I_{GDP}\ \ \widehat{KY}_{GDP}\ \ \Delta\widehat{Y}_{GDP}]$

9.15 Instability in the first-step VARs

I test the stability of the first-step VARs with two approaches[81]. In the first approach, the stability of the VAR is examined using a variety of tests including the Nyblom test, the MW test, the EW test and the QLR test. The MW, EW and QLR tests are based on the Chow Wald statistics. They are computed by excluding 30% of the observations at each end of the sample. The choice of a relatively large trimming

[79]The median-unbiased estimation method is used because the OLS estimator of the largest root is biased towards zero. The MUE estimates and their 90% confidence intervals are computed using the method described by Stock (1991).

[80]In the literature, the spline model has been used to approximate the trend component in unemployment or output. It is used in Staiger, Stock and Watson (1997a and 1997b) to model the NAIRU, and in Bernanke, Gertler and Watson (1997) to detrend the output.

[81]When there are both temporary and persistent time variations in the VAR coefficients, The two approaches used here will capture instability in a persistent manner. Temporary time variations, on the other hand, is controled by adopting heteroskedasticity-and-autocorrelation-robust estimation.

Table 13: Instability in the VAR

1. Interest Rate Excluded

	Nyblom		MW		EW		QLR	
Equation	p-value	$\widehat{\lambda}$	p-value	$\widehat{\lambda}$	p-value	$\widehat{\lambda}$	p-value	$\widehat{\lambda}$
Investment	0.00	5.12	0.00	3.40	0.00	3.84	0.00	3.90
Capital price	0.00	5.18	0.00	3.44	0.00	3.92	0.00	3.96
Capital share	0.00	4.34	0.01	2.70	0.00	4.00	0.00	4.03
Output	0.00	4.10	0.01	2.45	0.00	3.72	0.00	3.84

2. Interest Rate Included

	Nyblom		MW		EW		QLR	
Equation	p-value	$\widehat{\lambda}$	p-value	$\widehat{\lambda}$	p-value	$\widehat{\lambda}$	p-value	$\widehat{\lambda}$
Investment	0.00	6.27	0.00	4.43	0.00	5.38	0.00	5.47
Capital price	0.00	6.06	0.00	4.24	0.00	4.58	0.00	4.63
Capital share	0.00	5.91	0.00	4.09	0.00	5.99	0.00	6.07
Output	0.00	5.39	0.00	3.63	0.00	5.49	0.00	5.57
Interest rate	0.00	7.16	0.00	5.15	0.00	6.83	0.00	6.90

Note: p-values are computed by simulating the null distribution of the test statistics. The estimates of $\widehat{\lambda}$ are obtained by simulating the distribution of the test statistics under the TVP alternative. In all simulations, (i) the number of replications is 5,000; (ii) the number of the grid points used to approximate the integrals is 500.

at each end is based on the moderate sample size, $T = 140$ and the number of parameters in the VAR. There are 21 coefficients in each VAR equation. Even a 30% trimming implies in the smallest subsample, only 42 observations are used for 21 unknowns.

In the second approach, the amount of the time variation in the reduced-form VAR is quantified. To do so, I estimate the λ in (10), using the median-unbiased estimation proposed by Stock and Watson (1998)[82]. A zero λ indicates no time variation in the VAR; A non-zero λ signifies the presence of instability; and a large value of λ corresponds to large amount of instability.

It turns out both approaches indicate that *all* VAR specifications used in the exercise exhibit parameter instability. For illustration purpose, I present the stability results for one VAR, which is

[82] The general idea of the median-unbiased estimation method is to explore the fact that the distribution of the stability tests under the TVP alternative (9) and (10) depends on λ and Σ. Hence, by appropriate normalization so that Σ is consistently estimable, the distribution depends on λ only. Then the 50^{th} quantile of the distribution can be inverted to obtain a median-unbiased estimator of λ. In the current application, I impose the normalization $\Sigma = E(Z_{t-1}Z'_{t-1})^{-1}\text{Var}(\varepsilon_t Z_{t-1})E(Z_{t-1}Z'_{t-1})^{-1}$.

a first-step model for the third investment model in (36). Using the notation of Table 12, this is a five-variable VAR in $\Delta\widehat{I}_{NFB}$, $\Delta\widehat{p}^I_{NFB}$, $\widehat{KY}_{NFB}$, Δi and $\Delta\widehat{Y}_{NFB}$. The number of lags of this VAR is four. Table 13 summarizes the testing results for this VAR on equation-by-equation basis. The null of stability is rejected in all equations. The median-unbiased estimates are all small but non-zero, ranging from 3.63 to 7.16. Thus the empirical evidence suggests the asymptotic nesting (10) is appropriate.

9.16 Inclusion / Exclusion of Interest Rate

Table 14: Granger Causality Test of a First-Step VAR

(1) Interest Rate Excluded

	Excluded variable			
	$\Delta\widehat{I}$	$\Delta\widehat{p}^I$	$\widehat{KY}$	$\Delta\widehat{Y}$
$\Delta\widehat{I}$ equation	$\surd$			
$\Delta\widehat{p}^I$ equation	$\surd$	$\surd$	$\surd$	$\surd$
$\widehat{KY}$ equation	$\surd$	$\surd$	$\surd$	$\surd$
$\Delta\widehat{Y}$ equation	$\surd$	$\surd$	$\surd$	$\surd$

(2) Interest Rate Included

	Excluded variable				
	$\Delta\widehat{I}$	$\Delta\widehat{p}^I$	$\widehat{KY}$	$\Delta\widehat{Y}$	Δi
$\Delta\widehat{I}$ equation	$\surd$				$\surd$
$\Delta\widehat{p}^I$ equation	$\surd$	$\surd$		$\surd$	$\surd$
$\widehat{KY}$ equation	$\surd$		$\surd$	$\surd$	$\surd$
$\Delta\widehat{Y}$ equation	$\surd$		$\surd$	$\surd$	$\surd$
Δi equation			$\surd$	$\surd$	$\surd$

Note: "$\surd$" means the p value of the Granger causality test of excluding a variable from the VAR is greater than 5%.

9.17 Weak Identification in Time-varying 2SLS Models (Section 7)

Consider the following linear IV regression model in which all regressors are endogenous and there are no included exogenous variables,

$$y_t = x_t'\beta_t + \varepsilon_t \tag{79}$$

where x_t is a $k \times 1$ vector of regressors, ε_t is a mean zero scalar disturbance term, β_t is a $k \times 1$ vector of time-varying coefficients which follows the TVP process $\beta_t - \beta_{t-1} = \tau v_t$ with $\tau = \lambda_\beta/T$ and $v_t \sim N(0, \Sigma_\beta)$. The regressors, x_t, are connected to the set of instruments, z_t, by

$$x_t' = z_t'\Pi + u_t' \tag{80}$$

under β_0, where z_t is a $d \times 1$ vector of instruments, Π is a $d \times k$ coefficient matrix, u_t is a $k \times 1$ vector of mean zero disturbances which satisfies $E(z_t u_t') = 0$; Note that, to get the asymptotic results, as long as contiguity holds, I only require the regressors and instruments maintain a stable relationship in the hypothetical case of x_t being generated by the stable regression $y_t = x_t'\beta_0 + \varepsilon_t$.

The weak idendification, i.e., the weak correlation between the regressors, x_t, and the instruments, z_t, is modeled, following Staiger and Stock (1997), as

$$\Pi = T^{-1}\Gamma, \tag{81}$$

from which it is easy to show $T^{-1} \xrightarrow{p} 0$. In addition, the vector of instruments, z_t, and disturbances $\{\varepsilon_t, u_t\}$ satisfy the following assumptions.

- Assumption B1: $\{z_t, \varepsilon_t, u_t\}$ *follows a stationary process under* β_0 .
- Assumption B2: $z_t\varepsilon_t$ *and* $z_t u_t$ *are martingale difference sequences under* β_0 *with* $E(z_t z_t'\varepsilon_t^2) = \Sigma_{z\varepsilon}$ *and* $E(z_t z_t' u_t^2) = \Sigma_{zu}$, *with* $\Sigma_{z\varepsilon}$ *and* Σ_{zu} *being positive definite.*
- Assumption B3: $E(z_t z_t') = \Sigma_{zz}$ *with* Σ_{zz} *being positive definite.*
- Assumption B4: *The sequence of densities of data under* β_t *is contiguous of the sequences of densities of the data under* β_0.

The estabilishment of contiguity is identical to the proof of contiguity in Appendix 9.1, and hence is not repeated. Then we have the following proposition. The proposition shows that the asymptotic distribution of the standard 2SLS estimator does not depend on $\{\lambda_\beta, \Sigma_\beta\}$, the parameters governing the β_t process.

Proposition: *Consider the time-varying parameter 2SLS regression defined in (79), (80) and (81). Let* $W_{z\varepsilon}(\cdot)$ *and* $W_{zu}(\cdot)$ *be two standard Brownian motions associated with* $z_t\varepsilon_t$ *and* $z_t u_t'$. *Then, under Assumptions B1 to B4, the standard 2SLS estimator,* $\widehat{\beta}^{2SLS}$, *is asymptotically independent of the* β_t *process. The limiting distribution of the standard 2SLS estimator is given by*

$$\widehat{\beta}^{2SLS} - \beta_0 \Rightarrow F_D(W_{zu})^{-1} F_N(W_{z\varepsilon}, W_{zu})$$

$$\text{where} \quad F_D(W_{zu}) = \Gamma'\Sigma_{zz}\Gamma + \Gamma'\Sigma_{zu}^{1/2}W_{zu}(1) + W_{zu}'(1)\Sigma_{zu}^{1/2}\Gamma + W_{zu}'(1)\Sigma_{zu}^{1/2}\Sigma_{zz}^{-1}\Sigma_{zu}^{1/2}W_{zu}(1)$$

$$F_N(W_{z\varepsilon}, W_{zu}) = \Gamma'\Sigma_{z\varepsilon}^{1/2}W_{z\varepsilon}(1) + W_{zu}'(1)\Sigma_{zu}^{1/2}\Sigma_{zz}^{-1}\Sigma_{z\varepsilon}^{1/2}W_{z\varepsilon}(1)$$

Proof: First, by Assumption B1 to B4, the following results can be established on some partial sums *under* β_t.

(i) $T^{-1/2}\sum_{t=1}^{[sT]} z_t u_t'(\beta_t - \beta_0) \Rightarrow \lambda_\beta E(z_t u_t')\Sigma_\beta^{1/2}\int_0^s W_\beta(r)dr = 0$
(ii) $T^{-1/2}\sum_{t=1}^{[sT]} z_t z_t'\Gamma(\beta_t - \beta_0) \Rightarrow \lambda_\beta \Sigma_{zz}\Gamma\Sigma_\beta^{1/2}\int_0^s W_\beta(r)dr$
(iii) $T^{-1/2}\sum_{t=1}^{[sT]} z_t\varepsilon_t \Rightarrow \Sigma_{z\varepsilon}^{1/2}W_{z\varepsilon}(s)$
(iv) $T^{-1/2}\sum_{t=1}^{[sT]} z_t u_t' \Rightarrow \Sigma_{zu}^{1/2}W_{zu}(s)$
(v) $T^{-1/2}\sum_{t=1}^{[sT]} z_t z_t \xrightarrow{p} \Sigma_{zz}$

The proof of result (i) to result (v) is similar to the proof of Lemma 1 in appendix 9.3, and hence is not repeated. To faciliate the analysis in matrix form , let $Z = [z_1, ..., z_T]'$, and X, ε, u are defined likewise; let $\widetilde{Z} = diag(z_1', ...z_T')$, and $\widetilde{X}$ and $\widetilde{V}$ are defined likewise; let $\widetilde{\beta} = [(\beta_1 - \beta_0)', ..., (\beta_T - \beta_0)']'$.

Let $P_z = Z(Z'Z)^{-1}Z'$. Then the 2SLS estimator of β_0, expressed in matrix form, is $\widehat{\beta}^{2SLS} = [X'P_zX]^{-1}X'P_zy$, from which,

$$\widehat{\beta}^{2SLS} - \beta_0 = [X'P_zX]^{-1}[X'P_z\varepsilon + X'P_z\widetilde{X}\widetilde{\beta}] \tag{82}$$

where the effects of the instability in β_t is captured by the omitted variable $X'P_z\widetilde{X}\widetilde{\beta}$. In what follows, I derive the limits of the three terms on the right-hand side of (82). The term $X'P_zX$ can be decomposed as

$$X'P_zX = (T^{-1/2}Z\Gamma + u)'P_z(T^{-1/2}Z\Gamma + u) = A_{1T} + A_{2T} + A_{3T} + A_{4T} \text{ with}$$

$$\begin{aligned}
A_{1T} &= T^{-1}\Gamma'Z'Z\Gamma = \Gamma'T^{-1}\sum z_t z_t'\Gamma \xrightarrow{p} \Gamma'\Sigma_{zz}\Gamma \\
A_{2T} &= T^{-1/2}\Gamma'Z'u = \Gamma'T^{-1/2}\sum z_t u_t' \Rightarrow \Gamma'\Sigma_{zu}^{1/2}W_{zu}(1) \\
A_{3T} &= T^{-1/2}u'Z\Gamma = A_{2T}' \Rightarrow W_{zu}'(1)\Sigma_{zu}^{1/2}\Gamma \\
A_{4T} &= u'P_zu = [T^{-1/2}\sum u_t z_t'][T^{-1}\sum z_t z_t']^{-1}[T^{-1/2}\sum z_t u_t'] \\
&\Rightarrow W_{zu}'(1)\Sigma_{zu}^{1/2}\Sigma_{zz}^{-1}\Sigma_{zu}^{1/2}W_{zu}(1)
\end{aligned}$$

where the limit of A_{1T} follows from result (v); limits of A_{2T} and A_{3T} follow from result (iv); and the limit of A_{4T} follows from results (iv) and (v). Similarly, the term $X'P_z\varepsilon$ can be computed as

$$X'P_z\varepsilon = (T^{-1/2}Z\Gamma + u)'P_z\varepsilon = B_{1T} + B_{2T} \text{ where}$$

$$\begin{aligned}
B_{1T} &= T^{-1/2}\Gamma'Z'\varepsilon = \Gamma'T^{-1/2}\sum z_t\varepsilon_t \Rightarrow \Gamma'\Sigma_{z\varepsilon}^{1/2}W_{z\varepsilon}(1) \\
B_{2T} &= u'P_z\varepsilon = [T^{-1/2}\sum u_t z_t'][T^{-1}\sum z_t z_t']^{-1}[T^{-1/2}\sum z_t\varepsilon_t] \\
&\Rightarrow W_{zu}'(1)\Sigma_{zu}^{1/2}\Sigma_{zz}^{-1}\Sigma_{z\varepsilon}^{1/2}W_{z\varepsilon}(1)
\end{aligned}$$

where the limit of B_{1T} follows from result (iii); and the limit of B_{2T} follows from results (iii), (iv) and (v). Finally consider the term $X'P_z\widetilde{X}\widetilde{\beta}$,

$$X'P_z\widetilde{X}\widetilde{\beta} = (T^{-1/2}Z\Gamma + u)'P_z(T^{-1/2}\widetilde{Z}\otimes\Gamma + \widetilde{u})\widetilde{\beta} = C_{1T} + C_{2T} + C_{3T} + C_{4T} \text{ where}$$

$$\begin{aligned}
C_{1T} &= T^{-1/2}\Gamma'Z'\widetilde{u}\widetilde{\beta} = \Gamma'T^{-1/2}\sum z_t u_t(\beta_t - \beta_0) \to 0 \\
C_{2T} &= u'P_z\widetilde{u}\widetilde{\beta} = [T^{-1/2}\sum u_t z_t'][T^{-1}\sum z_t z_t']^{-1}[T^{-1/2}\sum z_t u_t(\beta_t - \beta_0)] \to 0 \\
C_{3T} &= T^{-1}\Gamma'Z'(\widetilde{Z}\otimes\Gamma)\widetilde{\beta} = T^{-1/2}[T^{-1/2}\sum z_t z_t'\Gamma(\beta_t - \beta_0)] = O_p(T^{-1/2}) \to 0 \\
C_{4T} &= T^{-1/2}u'P_z(\widetilde{Z}\otimes\Gamma)\widetilde{\beta} \\
&T^{-1/2}[T^{-1/2}\sum u_t z_t'][T^{-1}\sum z_t z_t']^{-1}[T^{-1/2}\sum z_t z_t'\Gamma(\beta_t - \beta_0)] = O_p(T^{-1/2}) \to 0
\end{aligned}$$

where the limit of C_{1T} follows from result (i); the limit of C_{2T} follows from results (i), (iv) and (v); the limit of C_{3T} follows from result (ii) and the limit of C_{4T} follows from result (ii), (iv) and (v).

Summerizing, the limits of the three right-hand side terms in (82) are

$$\begin{aligned}
X'P_zX &\Rightarrow \Gamma'\Sigma_{zz}\Gamma + \Gamma'\Sigma_{zu}^{1/2}W_{zu}(1) + W_{zu}'(1)\Sigma_{zu}^{1/2}\Gamma + W_{zu}'(1)\Sigma_{zu}^{1/2}\Sigma_{zz}^{-1}\Sigma_{zu}^{1/2}W_{zu}(1) \\
X'P_z\varepsilon &\Rightarrow \Gamma'\Sigma_{z\varepsilon}^{1/2}W_{z\varepsilon}(1) + W_{zu}'(1)\Sigma_{zu}^{1/2}\Sigma_{zz}^{-1}\Sigma_{z\varepsilon}^{1/2}W_{z\varepsilon}(1) \\
X'P_z\widetilde{X}\widetilde{\beta} &\xrightarrow{p} 0,
\end{aligned}$$

from which, it is evident the effects of the ignored time variation in β_t is negligible asymptotically. Substituting the above limits to (82) gives the limting distribution of $\widehat{\beta}^{2SLS}$, which is free of λ_β, Σ_β, and $W_\beta(\cdot)$ − the parameters and Brownian motion associated with the β_t process.