Weak Instruments Estimation and Inference in the Presence of Parameter Instability*

Hong Li[†]
Brandeis University

Zhijie Xiao[‡]
Boston College

Abstract

The paper considers time series models when (i) some or all of the parameters are weakly identified and (ii) some or all of the parameters are time varying. Following the previous literature, the weak correlations between the instruments and the relevant first order condition are modeled as local to zero, and we focus on an empirically relevant case with moderately large instabilities, which are well approximated by a local asymptotic device. We show that for many forms of the instability and a large class of econometric models, weak instruments effect dominates in the simultaneous presence of weak identification and parameter instability. As a result, when making inference for the weakly identified parameters, the weak identification robust methods in the literature remain valid. One can thus ignore moderate instabilities in the model and still obtain approximately correct inference for the weakly identified parameters. Monte Carlo experiments indicate that for moderate size of parameter instability, asymptotic theory developed in this paper provides good empirical guidance.

JEL Classification: C32

Keywords: Weak Instrument, Parameter Stability

*We thank Mark Watson, Zhongjun Qu, Ulrich Müller and participants of 2007 Econometrics Society Summer meeting for helpful comments. All errors are ours.

[†]Department of Economics, Brandeis University, Waltham MA 02454 USA, tel: 781-736-5007, fax: 781-736-2269, email: hli@brandeis.edu.

[‡]Department of Economics, Boston College, Chestnut Hill, MA 02467 USA, tel: 617-552-1709, fax: 617-552-2308, email: zhijie.xiao@bc.edu.

1 Introduction

Many macroeconomic and finance problems often confront two issues: weak identification and structural instability. Either of the two issues may invalidate the conventional econometric inference that ignores it. For this reason, weak identification has received increased attention, and there has traditionally been a long-standing research effect in structural instability.

Weak identification arises when the postulated distribution of the sample data is very similar for different values of the parameter, hence it is difficult for the researcher to distinguish between alternative values of the parameter. It is known that weak identification is one source of the poor approximation of asymptotic normality to the finite sample behavior of estimators and statistics, see for instance Stock, Wright and Yogo (2002) and Stock and Yogo (2005a,b). Weak identification has been shown to be a concern in the estimation of macroeconomic equations with expectations, as pointed out by Ma (2002), Mavroeidis (2004, 2005), Fuhrer and Rudebusch (2004) and Dufour, Khalaf and Kichian (2006). Seminal papers of Staiger and Stock (1997) and Stock and Wright (2000) develop weak instrument asymptotics for linear and nonlinear econometric models. They show that limits under the new asymptotic device provide a better approximation for the finite sample behavior of estimators. Stock and Wright (2000) and Kleibergen (2005) provide statistics which can be used to construct weak-instrument robust inference for the parameters.

Structural instability can be interpreted as parameter values that evolve over time. Parameter instability is an empirically widespread phenomenon, see Ghysels (1998), Stock and Watson (1998), Boivin (1999) and Cogley and Sargent (2005). Parameter instability is typically caused by changes in economic environment and policies, which in turn induce behavioral changes of economic agents and are reflected in time-varying parameters in many time series econometric relationships. In addition, misspecification of econometric models can also manifest themselves in the form of time-varying parameters. Ignoring parameter instabilities that are presented in the econometric models may invalidate the standard asymptotic inference. Econometric theory has focused on the problem of stability testing, see Nyblom (1989), Andrews (1993, 2003), Andrews and Ploberger (1994), Sowell (1996), Bai, Lumsdaine and Stock (1998), Hansen (2000), Elliott and Müller (2005), just to name a few. Some work has also addressed estimation and inference in the presence of parameter instability, see for instance Ghysels and Hall (1990a,b), Li (2008) and Li and Müller (2008).

In practice, an important empirical issue is that parameter instability and weak identification may occur simultaneously. It is easy to conceive of many applications with such a dual problem.

[*Example 1*]: Consider a forward-looking version of Taylor's (1993) rule.

$$i_t = \alpha_t + \beta_t E_t \pi_{t+1} + \gamma_t E_t x_{t+1} + \rho_t i_{t-1} + \varepsilon_t \tag{1}$$

where i is a short-term interest rate, π and x are inflation and output gap. E_t denotes rational expectations conditional on the information set at date t. (1) is a structural characterization of monetary policy conduct. The parameter instability indicates the regime shift due to the change of monetary policy over time. This type of models with forward-looking feature often suffer weak identification. If coefficients in (1) are all constant, then statistics in Stock and Wright (2000) and Kleibergen (2005) can be used to conduct weak-instrument robust inference. But here, the additional issue of unstable coefficients complicates the problem. Such a case naturally creates a situation which involves both structural instability and weak identification.

[*Example 2*]: Consider a three equation system consisting of

$$x_t = \alpha E_t x_{t+1} + \beta(i_t - E_t \pi_{t+1}) + \varepsilon_t \tag{2}$$

$$\pi_t = \rho_{1t}(L)\pi_{t-1} + \rho_{2t}(L) i_{t-1} + \rho_{3t}(L) y_{t-1} + \eta_{\pi,t} \tag{3}$$

$$i_t = \gamma_{1t}(L)\pi_{t-1} + \gamma_{2t}(L) i_{t-1} + \gamma_{3t}(L) y_{t-1} + \eta_{i,t} \tag{4}$$

where (2) is an Euler equation for output gap x. Again, the forward-looking nature of (2) makes it vulnerable to weak identification. But (2) is assumed to be stable because its coefficients reflect private agents' preferences. On the other hand, the evolution of inflation π and nominal interest rate i is described by reduced-form equations (3) and (4) which are arguably strongly identified. Moreover, the two reduced-form equations are unstable because (i) rational agents adapt their optimal behavior to policy changes (equation 3); (ii) monetary policy varies over time (equation 4). Thus the three-equation system provides an example where the model is partially unstable and partially weakly identified.

Applications such as the two examples given above call for methods that address both parameter instability and weak identification simultaneously. But so far the two strands of research to a large extent address the two issues separately - the literature on weak instruments has seldom considered the possibility of the additional data problem of structural instability, and vice versa[1]. A natural question to ask is: how does one conduct valid inference in an unstable model with weak identification? One possible approach would be to estimate the break dates, then standard weak-identification robust inference can be applied to coefficients in various regimes. But to date no method is available in the literature to estimate the break dates in the presence of weak instruments.

Therefore, in this paper we approach the issue in an alternative way. We ask the following questions: If we are concerned about the average values of the time-varying parameters, then (i) what is the consequence of the conventional estimation methods that ignore the parameter instability? (ii) could the standard weak-instrument robust procedures in the literature remain valid?

In the literature, this paper is closely related to Staiger and Stock (1997), Stock and Wright (2000), Ghysels and Hall (1990a,b), Li (2008) and Li and Müller (2008). Staiger and Stock (1997) and Stock and Wright (2000) study inference in stable models under weak identification. Ghysels and Hall (1990a,b), Li (2008) and Li and Müller (2008) conduct analysis on inference in unstable models with strong identification. This paper, to fill up the gap in the literature, focuses on inference in unstable econometric models with weak identification.

Looking ahead, the main findings of the paper can be summarized as follows.

1. With the simultaneous presence of weak identification and parameter instability, the distortion on inference (that is, the departure from the conventional distribution) is dominated by weak identification.

2. In an unstable model with weak identification only (that is, there is no strongly identified component in the model), weak-instrument robust statistics proposed in the literature, such as those in Stock and Wright (2000) and Kleibergen (2005), can still be used to construct valid inference for the average values of the unstable parameters.

3. In an unstable model with both weak and strong identifications, weak-instrument ro-

[1]The only exception is the recent paper of Caner (2007) who studies stability testing with unknown break point under weak identification. Our paper, instead, focuses on inference problem.

bust methods such as those in Stock and Wright (2000) and Kleibergen (2005) can be used to make valid inference for the weakly identified component, after the strongly identified component is concentrated out. But unlike the case of stable parameter models, these statistics can not be used to construct inference for the strongly identified coefficients as long as the strongly identified coefficients are time-varying.

To illustrate how to apply these results, consider the time-varying Taylor rule in Example 1. (1) can be estimated by (i) ignoring the parameter instability and (ii) using lagged endogenous variables as instruments. If all parameters are weakly identified, valid confidence sets of the average values of α_t, β_t, γ_t and ρ_t paths can be constructed using the Stock-Wright and Kleibergen statistics. On the other hand, if only a subset of regressors, say the two expected future variables, are weakly correlated with the instruments, then our results suggest that the Stock-Wright and Kleibergen statistics after concentrating out α and ρ can be used to construct valid inference for the average values of β_t and γ_t.

Similarly, suppose the three-equation system in Example 2 is jointly estimated by conventional methods. Then according to our results, after coefficients in (3) and (4) (the two strongly-identified equations) are concentrated out, Stock-Wright and Kleibergen statistics can be used to construct valid inference for α and β in the Euler equation (2). In a more general case where the weakly-identified parameters are also unstable, the resulting confidence set is interpreted as the confidence set for the average values of α and β paths.

Of course, in applications of our theoretical results, one must decide whether these asymptotic considerations yield accurate approximations in small samples. Our Monte Carlo study shows that, for the size of parameter instability typically encountered in practice, our asymptotic results provide a reasonable guidance. But the asymptotic theory breaks down in finite samples if the instability is very large in magnitude.

The rest of the paper is organized as follows. Section 2 studies the leading case of single-equation linear regression models with weak instruments and parameter instability. Section 3 generalizes the results of Section 2 to nonlinear GMM models. Section 4 presents two Monte Carlo experiments. Section 5 concludes. Mathematical details are collected in the appendix.

2 Single-Equation Linear Models

In this section, we focus on the leading case of single-equation linear regressions and study the impact of parameter instability on weak instrument estimation. Analysis conducted in this section combines the literatures on weak instrument and parameter instability.

2.1 The model and assumptions

The structural model of interest is

$$y_t \;=\; \alpha_t' Y_{At} + \beta_t' Y_{Bt} + u_t \tag{5}$$

where y_t is a scalar dependent variable, $\theta_t = [\alpha_t' \; \beta_t']'$ is a vector of (possibly) time-varying parameters, and $Y_t = [Y_{At}' \; Y_{Bt}']'$ is an $n \times 1$ vector of endogenous variables[2]. Y_{At} and Y_{Bt} are $n_A \times 1$ and $n_B \times 1$ respectively with $n = n_A + n_B$. Y_{At} is weakly correlated with the instruments and Y_{Bt} is strongly correlated with the instruments. Let the reduced forms of Y_{At} and Y_{Bt} be

$$\begin{aligned} Y_{At} &= \Pi_A' Z_t + V_{At} \\ Y_{Bt} &= \Pi_B' Z_t + V_{Bt} \end{aligned} \tag{6}$$

respectively where Z_t denotes the $p \times 1$ vector of instruments. The strength of the link between the endogenous variables and the instruments is measured by Π_A and Π_B.[3] Let $U_t = [u_t \; V_A' \; V_B']'$. Following the convention in the literature of weak instruments, we have the following assumption.

Assumption 2.A : *The error U_t satisfies the following conditions:*

(a) U_t is mean zero, serially uncorrelated and homoskedastic.

(b) The covariance matrix of U_t, denoted by Σ_U, is partitioned so that $Eu_t^2 = \sigma_{uu}$,

[2] We assume for notational convenience that the structural equation of interest contains no exogenous variables. This is readily relaxed using the standard projection arguments.

[3] Π_A and Π_B are assumed to be time-invariant. But making Π_A and Π_B time-varying does not change the result. Thus to focus on the instability of the coefficient of interest, α_t and β_t, we will keep Π_A and Π_B constant throughout the paper.

$Eu_tV_{At} = \Sigma_{uV_A}$, $Eu_tV_{Bt} = \Sigma_{uV_B}$, $EV_{At}V'_{At} = \Sigma_{V_AV_A}$, $EV_{Bt}V'_{Bt} = \Sigma_{V_BV_B}$,

$EV_{At}V'_{Bt} = \Sigma_{V_AV_B}$.

(c) $E(Z_tU_t) = 0$.

Unstable parameters may vary in either stochastic or nonstochastic manners. Take α_t as an example. Suppose the instability is *stochastic*, then following Nyblom (1989), Stock and Watson (1998), Elliot and Müller (2006) and Li (2008), the instability is modeled as

$$\alpha_t = \alpha_{t-1} + v_t, \quad \text{with } v_t = w_t/T \tag{7}$$

where w_t is a mean zero disturbance term. This modeling strategy is supported by empirical studies such as Stock and Watson (1996), who find that instability presented in most macroeconomic relations is characterized by persistent but small changes in the coefficients[4]. Note that (7) can be written as

$$\alpha_t = \alpha_0 + T^{-1/2}\left[T^{-1/2}\sum_{i=1}^{t} w_i\right] \tag{8}$$

where $T^{-1/2}\sum_{i=1}^{t} w_i$ obeys a Functional Central Limit Theorem and converges to a Wiener process under general regularity conditions. α_0 is the initial value of the α sequence. It can also be understood as the average value of the $\{\alpha_t\}$ sequence because it is easy to verify that $T^{-1}\sum_{t=1}^{T} \alpha_t$ converges in probability to α_0.

Suppose, on the other hand, the instability follows a *nonstochastic* path, then following the modeling strategy in Andrews and Ploberger (1994), Sowell (1996) and Li and Müller (2008), it can be modeled as

$$\alpha_t \;=\; \alpha_0 + T^{-1/2}g(t/T) \tag{9}$$

where $g(t/T)$ is a bounded nonrandom function that is piece-wise continuous with a finite number of discontinuities. For instance, single or multiple discrete break(s) that are frequently studied in the literature belong to this category. As in the case of stochastic instabilities, the initial value α_0 can be interpreted as the average value of the $\{\alpha_t\}$ sequence.

[4]The random walk specification in (7) captures the persistence in the time-varying parameters. The local-to-zero device in $v_t = w_t/T$ formalizes the small period-to-period variations.

Given the discussion above, for notational convenience, the stochastic instability of (8) and the nonstochastic instability of (9) are summarized in a unified framework. The instability in β_t is handled similarly.

Assumption 2.B : *The time-varying parameter vectors α_t and β_t in (5) satisfy*

$$\alpha_t = \alpha_0 + T^{-1/2} W_{\alpha,T}(t/T) \quad \text{and} \quad \beta_t = \beta_0 + T^{-1/2} W_{\beta,T}(t/T)$$

where $W_{\alpha,T}(t/T)$ and $W_{\beta,T}(t/T)$ can be either $O_p(1)$ random functions that converge to Wiener processes or $O(1)$ nonrandom functions that are bounded, piece-wise continuous and with a finite number of discontinuities. Moreover, let $W_\alpha(s)$ and $W_\beta(s)$ be the limits of $W_{\alpha,T}(s)$ and $W_{\beta,T}(s)$ respectively for any $s \in [0, 1]$.

In other words, following (8) and (9), $W_{\alpha,T}(t/T) = T^{-1/2} \sum_{s=1}^{t} w_s$ for stochastic instability and $W_{\alpha,T}(t/T) = g(t/T)$ for non-stochastic instability. So does the instability in β_t. Moreover, some elements of $W_{\alpha,T}(t/T)$ and $W_{\beta,T}(t/T)$ are allowed to be zero if the corresponding subsets of α_t and β_t are time invariant. Regardless of the nature of the unstable parameters, the magnitude of the instabilities is of order $T^{-1/2}$ in a sample of size T. This is the neighborhood in which efficient tests of parameter stability have nontrivial local asymptotic power. We emphasize that this does not mean that our results only apply to economically insignificant instabilities. Linde (2001) argues that economically important changes in monetary policy lead to parameter instabilities that are small in the sense of being difficult to detect empirically. More generally, Stock and Watson (1996) document that instabilities in macroeconomic data are often borderline significant. In such instances, accurate approximations are generated by a modeling strategy in which there is only limited information about the instability asymptotically, as in the $T^{-1/2}$ neighborhood.

Next, to differentiate weak identification from strong identification, Π_A is modeled to be local-to-zero, and Π_B is assumed to be fixed.

Assumption 2.C:

(a) $\Pi_A = \Pi_{AT} = C_A/\sqrt{T}$, *where C_A is an $p \times n_A$ matrix.*

(b) Π_B *is a fixed $p \times n_B$ matrix that does not change with T.*

The rate of $T^{-1/2}$ in Assumption 2.C part (a) yields tractable asymptotic approximations to the sampling distributions of conventional estimators and test statistics when the relevant parameters are weakly identified. This nesting reflects the fact that weak instruments lead to an objective function that is nearly flat in α.[5]

Note that the only fundamental difference between the weakly identified component $\alpha_t' Y_{At}$ and the strongly identified component $\beta_t' Y_{Bt}$ in equation (5) is the strength of link between the endogenous variables and the instruments. All other aspects are treated symmetrically.

Finally, rather than making primitive assumptions on the errors (u_t, V_{At} and V_{Bt}), and the instruments Z_t, we impose some moment requirements that the errors and instruments must satisfy. Let "$\xrightarrow{p}$" denote convergence in probability and "$\Rightarrow$" denote convergence in distribution.

Assumption 2.D: *The following limits hold jointly for the model of (5) and (6):*

(a) $\left(T^{-1}\sum_{t=1}^T u_t^2, T^{-1}\sum_{t=1}^T V_{At}u_t, T^{-1}\sum_{t=1}^T V_{Bt}u_t, T^{-1}\sum_{t=1}^T V_{At}V_{At}', T^{-1}\sum_{t=1}^T V_{Bt}V_{Bt}'\right) \xrightarrow{p}$

$(\sigma_{uu}, \Sigma_{V_A u}, \Sigma_{V_B u}, \Sigma_{V_A V_A}, \Sigma_{V_B V_B})$.

(b) $T^{-1}\sum_{t=1}^T Z_t Z_t' \xrightarrow{p} Q_{zz}$ *where* $Q_{zz} = EZ_tZ_t'$.

(c) $\left(T^{-1/2}\sum_{t=1}^T Z_t u_t, T^{-1/2}\sum_{t=1}^T Z_t V_{At}, T^{-1/2}\sum_{t=1}^T Z_t V_{Bt},\right) \Rightarrow (\Psi_{Zu}, \Psi_{ZV_A}, \Psi_{ZV_B})$

where $(\Psi_{Zu}, vec(\Psi_{ZV_A})' vec(\Psi_{ZV_B})')' = \mathcal{N}(0, \Sigma_U \otimes Q_{zz})$.

(d) $T^{-1}\sum_{t=1}^T x_{i,t}x_{j,t}' W_{k,T}(t/T) \Rightarrow E(x_{i,t}x_{j,t}')\int_0^1 W_k(s)ds$,

$T^{-1}\sum_{t=1}^T W_{k,T}(t/T)' x_{i,t}x_{j,t}' W_{k,T}(t/T) \Rightarrow \int_0^1 W_k(s)' E(x_{i,t}x_{j,t}') W_k(s)ds$,

where $k = \alpha, \beta$; $x_{i,t}$ *and* $x_{j,t}$ *can be any of* u_t, $V_{A,t}$, $V_{B,t}$ *and* Z_t.

Assumption 2.D holds under general weak primitive assumptions. For example, suppose

[5] An alternative way to understand the main idea behind this modeling strategy is that we are interested in the case where due to the weak correlation between the instrument Z and the endogenous regressor Y_A, the mean of the first-stage F statistics testing for $\Pi_A = 0$ is small even in large samples. In such instances, this local-to-zero Π_A asymptotics will make the F-statistic an $O_p(1)$ random variable with weak instruments, while the fixed Π_A asymptotics provide poor approximations by making the F-statistic explosive.

all errors are martingale difference sequence with four moments, the instruments are integrated of order zero with four moments and they satisfy some weak conditions regarding limiting dependence, then Assumption 2.D follows from the weak Law of Large Numbers and the Central Limit Theorem[6].

The main interest of this section is to investigate whether the standard inference procedures for linear regressions with weak instruments, assuming erroneously constant parameters, are possible to remain valid in the presence of parameter instability. To develop some intuition, a simple scalar example is considered next.

2.2 An illustrative example

Consider a scalar IV regression with a scalar instrument. (5) and (6) are simplified to

$$\begin{aligned} y_t &= \alpha_t Y_t + u_t \\ Y_t &= (C/\sqrt{T}) Z_t + V_t \end{aligned} \tag{10}$$

Note that the structural equation can be equivalently written as

$$y_t = \alpha_0 Y_t + [(\alpha_t - \alpha_0) Y_t + u_t] \tag{11}$$

where the neglected time variation is an omitted variable. When the conventional IV estimation is conducted by mistakenly ignoring the instability of α_t, there are two consequences: (i) the estimated α_0 is the average value of the α_t sequence; (ii) the omitted instability induces an additional source of endogeneity. The key issue is whether the endogeneity caused by the original error u_t dominates the additional endogeneity induced by the omitted instability.

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. The IV estimator $\widehat{\alpha} = (Z'Y)^{-1} Z'y$ can be rewritten as

$$\widehat{\alpha} = \alpha_0 + \frac{\sum Z_t u_t + \sum Z_t Y_t (\alpha_t - \alpha_0)}{\sum Z_t Y_t}. \tag{12}$$

The property of $\widehat{\alpha}$ is then determined by the asymptotic behavior of $\sum Z_t u_t$, $\sum Z_t Y_t (\alpha_t - \alpha_0)$ and $\sum Z_t Y_t$. Among the three terms, the strength of the instrument affects

[6]For brevity, the establishment of Assumption 2.D part (d) is skipped. Readers are referred to Stock and Watson (1998), Li (2008) and Li and Müller (2008) for the proof.

$\sum Z_t Y_t(\alpha_t - \alpha_0)$ and $\sum Z_t Y_t$ via the endogenous regressor Y_t. Also, $\sum Z_t Y_t(\alpha_t - \alpha_0)$ is the only term that captures the effect of the neglected parameter instability.

Under weak identification, we have

$$T^{-1/2}\sum Z_t u_t \ \Rightarrow \ \Psi_{Zu} \tag{13}$$
$$T^{-1/2}\sum Z_t Y_t(\alpha_t - \alpha_0) \ \xrightarrow{p} \ 0 \tag{14}$$
$$T^{-1/2}\sum Z_t Y_t \ \Rightarrow \ Q_{zz}C + \Psi_{ZV} \tag{15}$$

where the limits in (13) to (15) are derived in the appendix. According to (14), the term $T^{-1/2}\sum Z_t Y_t(\alpha_t - \alpha_0)$ is $o_p(1)$, which means the parameter instability has only negligible effect when the instrument is weak. So the limiting distribution of the IV estimator would not be affected by the instability in α_t. To see this more clearly, Substituting (13)−(15) in (12), we get

$$\widehat{\alpha} - \alpha_0 \ \Rightarrow \ \frac{\Psi_{Zu}}{Q_{zz}C + \Psi_{ZV}}. \tag{16}$$

Note that the distribution in (16) is *identical* to the limit one would obtain in a weakly identified model with no parameter instability, see Staiger and Stock (1997). By the same intuition, the weak instrument robust inference for α_0 will remain largely unaffected by the instability in α_t.

2.3 General asymptotic results

The intuition developed in the scalar example easily carries over to multiple regressions where all parameters are weakly identified. But what if only a subset of the parameters are weakly identified? In this section, we provide asymptotic analysis for the general case with both weak and strong identifications.

Consider the two-stage least square (TSLS) and limited information maximum likelihood (LIML) estimators that are most frequently used in applications. The TSLS and LIML estimators minimize

$$\begin{aligned} S_T^{TSLS}(\alpha,\beta) &= \left[T^{-1/2}Z'(y - Y_A\alpha - Y_B\beta)\right]' M_T^{TSLS}\left[T^{-1/2}Z'(y - Y_A\alpha - Y_B\beta)\right] \\ S_T^{LIML}(\alpha,\beta) &= \left[T^{-1/2}Z'(y - Y_A\alpha - Y_B\beta)\right]' M_T^{LIML}(\alpha,\beta)\left[T^{-1/2}Z'(y - Y_A\alpha - Y_B\beta)\right] \end{aligned} \tag{17}$$

respectively, where the weighting matrices are $M_T^{TSLS} = \left[T^{-1}Z'Z\right]^{-1}$ and $M_T^{LIML}(\alpha,\beta) =$

$[T^{-1}Z'Z]^{-1}\left[T^{-1}(y - Y_A\alpha - Y_B\beta)'(y - Y_A\alpha - Y_B\beta)\right]^{-1}$ with $T^{-1}(y - Y_A\alpha - Y_B\beta)'(y - Y_A\alpha - Y_B\beta)$ being an estimator of σ_{uu} evaluated at (α, β).

Our analysis will proceed by first presenting the asymptotic limits of S_T^{TSLS} and S_T^{LIML} for any given pair of (α, β). For the weakly identified component, define $\lambda_A = Q_{zz}^{1/2}C_A$, $\xi_{V_A} \sim \mathcal{N}(0, \Sigma_{V_AV_A})$ and $\zeta_A = \lambda_A + \xi_{V_A}$; for the strongly identified component, define $\lambda_B = Q_{zz}^{1/2}\Pi_B$ and $\xi_{V_B} \sim \mathcal{N}(0, \Sigma_{V_BV_B})$. Let $\beta = \beta_0 + b/\sqrt{T}$, so we consider local departure of β from β_0.

Proposition 1: *Consider the regression model described in (5) and (6). Under Assumptions 2.A to 2.D, the following limits hold jointly at* $(\alpha, \beta_0 + b/\sqrt{T})$.

(a) Scaled sample moment condition:

$$T^{-1/2}Q_{zz}^{-1/2}Z'(y - Y_A\alpha - Y_B\beta) \Rightarrow \xi_u - \zeta_A(\alpha - \alpha_0) - \lambda_B b + \lambda_B \int W_\beta(s)ds.$$

(b) TSLS objective function:

$$S_T^{TSLS}(\alpha, \beta_0 + b/\sqrt{T}) \Rightarrow S_*^{TSLS}(\alpha, b) \text{ with}$$

$$S_*^{TSLS}(\alpha, b) = \left[\xi_u - \zeta_A(\alpha - \alpha_0) - \lambda_B(b - \textstyle\int W_\beta(s)ds)\right]' \left[\xi_u - \zeta_A(\alpha - \alpha_0) - \lambda_B(b - \textstyle\int W_\beta(s)ds)\right].$$

(c) LIML objective function:

$$S_T^{LIML}(\alpha, \beta_0 + b/\sqrt{T}) \Rightarrow S_*^{LIML}(\alpha, b) \text{ with}$$

$$S_*^{LIML}(\alpha, b) = \frac{\left[\xi_u - \zeta_A(\alpha - \alpha_0) - \lambda_B(b - \int W_\beta(s)ds)\right]' \left[\xi_u - \zeta_A(\alpha - \alpha_0) - \lambda_B(b - \int W_\beta(s)ds)\right]}{\sigma_{uu} + (\alpha - \alpha_0)'\Sigma_{V_AV_A}(\alpha - \alpha_0) - 2\Sigma_{uV_A}(\alpha - \alpha_0)}.$$

Proofs of all propositions are given in the appendix.

Proposition 1 says that for an arbitrary pair of (α, β), the departure of the scaled moment condition and objective functions from the conventional distributions is due to three sources: (i) deviation of α from α_0, (ii) deviation of β from β_0, and (iii) β_t, the parameter instability from the strongly identified part of the model. Among them, source (i) is related to the weakly identified component, while sources (ii) and (iii) are related to the strongly identified component.

Next, we derive the limiting distribution of the TSLS and LIML estimators by minimizing the limiting objective functions in Proposition 1 parts (c) and (d). Optimization is conducted sequentially. β, the strongly identified coefficient vector, is concentrated out. Then minimizing the concentrated limiting objective function delivers the asymptotic distribution of α. Let $S_T(\alpha, \widehat{\beta}(\alpha))$ be the concentrated objective functions and $M_{\lambda_B} = I - \lambda_B(\lambda_B'\lambda_B)^{-1}\lambda_B'$.

Proposition 2: *Consider the regression model described in (5) and (6). Under Assumptions 2.A to 2.D, the following limits hold jointly.*

(a) Scaled concentrated sample moment condition:

$$T^{-1/2}Q_{zz}^{-1/2}Z'(y - Y_A\alpha - Y_B\widehat{\beta}(\alpha)) \Rightarrow M_{\lambda_B}\left[\xi_u - \zeta_A(\alpha - \alpha_0)\right].$$

(b) Concentrated objective functions:

$$S_T^{TSLS}(\alpha, \widehat{\beta}(\alpha)) \Rightarrow S_*^{TSLS}(\alpha) = \left[\xi_u - \zeta_A(\alpha - \alpha_0)\right]' M_{\lambda_B}\left[\xi_u - \zeta_A(\alpha - \alpha_0)\right],$$

$$S_T^{LIML}(\alpha, \widehat{\beta}(\alpha)) \Rightarrow S_*^{LIML}(\alpha) = \frac{\left[\xi_u - \zeta_A(\alpha - \alpha_0)\right]' M_{\lambda_B}\left[\xi_u - \zeta_A(\alpha - \alpha_0)\right]}{\sigma_{uu} + (\alpha - \alpha_0)'\Sigma_{V_AV_A}(\alpha - \alpha_0) - 2\Sigma_{uV_A}(\alpha - \alpha_0)}.$$

(c) TSLS estimator:

$\left(\widehat{\alpha}^{TSLS}, \sqrt{T}(\widehat{\beta}^{TSLS} - \beta_0)\right) \Rightarrow (\alpha_*^{TSLS}, b_*^{TSLS})$ *where*

$\alpha_*^{TSLS} = \arg\min_{\alpha\in\mathcal{A}} S_*^{TSLS}(\alpha) = \alpha_0 + (\zeta_A' M_{\lambda_B}\zeta_A)^{-1}\zeta_A' M_{\lambda_B}\xi_u,$

$b_*^{TSLS} = (\lambda_B'\lambda_B)^{-1}\lambda_B'\left[\xi_u - \zeta_A(\alpha_*^{TSLS} - \alpha_0)\right] + \int W_\beta(s)ds.$

(d) LIML estimator:

$\left(\widehat{\alpha}^{LIML}, \sqrt{T}(\widehat{\beta}^{LIML} - \beta_0)\right) \Rightarrow (\alpha_*^{LIML}, b_*^{LIML})$ *with*

$\alpha_*^{LIML} = \arg\min_{\alpha\in\mathcal{A}} S_*^{LIML}(\alpha),$

$b_*^{LIML} = (\lambda_B'\lambda_B)^{-1}\lambda_B'\left[\xi_u - \zeta_A(\alpha_*^{LIML} - \alpha_0)\right] + \int W_\beta(s)ds.$

Comments:

1. Effects of parameter instabilities on estimators

The estimator of the weakly identified coefficients, $\widehat{\alpha}$, has a nonstandard distribution, which is *solely* caused by weak instruments. Instability arising from any parts in the model does not play a role. The estimator of the strongly identified coefficients, $\widehat{\beta}$, also has a nonstandard distribution. But the nonstandard distribution of $\widehat{\beta}$ is due to the *joint* effect of the weak instruments and the ignored instability in the strongly identified coefficients themselves.

2. Concentrated moment function and objective functions

Recall in Proposition 1, the scaled sample moment condition, and hence the objective functions evaluated at arbitrary (α, β) are affected by the β-instability. Proposition 2 shows, however, that once the strongly identified coefficients are concentrated out, the resulting concentrated sample moment condition and objective functions are independent of the β-instability. The reason is that, β-instability affects the concentrated moment condition via two channels: (i) the direct impact captured by $\lambda_B \int W_\beta(s)ds$ (see Proposition 1(a)), and (ii) the indirect impact embedded in the optimal solution for $\widehat{\beta}$ (see Proposition 2 parts (c) and (d)). The two effects operate in opposite directions with identical magnitude, and hence are canceled out.

3. The case where all parameters are weakly identified

If all parameters are weakly identified, like the scalar example discussed in Section 2.2, then the model in (5) and (6) can be modified by (i) setting $\beta_t = 0$ for any t in (5) and (ii) shutting down the second equation in (6). Accordingly, asymptotic results for these regressions can be obtained by shutting down the strongly identified components in Proposition 2, which gives the following corollary.

Corollary 1: *Consider a regression model where all parameters are weakly identified. Under Assumptions 2.A to 2.D, the following limits hold jointly.*

(a) TSLS objective function and estimator:

$$S_T^{TSLS}(\alpha) \Rightarrow S_*^{TSLS}(\alpha) = [\xi_u - \zeta_A(\alpha - \alpha_0)]' [\xi_u - \zeta_A(\alpha - \alpha_0)]$$

$$\widehat{\alpha}^{TSLS} \ \Rightarrow \ \alpha_*^{TSLS} = \arg\min_{\alpha \in \mathcal{A}} S_*^{TSLS}(\alpha) = \alpha_0 + (\zeta_A' \zeta_A)^{-1} \zeta_A' \xi_u$$

(b) LIML objective function and estimator:

$$S_T^{LIML}(\alpha) \Rightarrow S_*^{LIML}(\alpha) = \frac{[\xi_u - \zeta_A(\alpha - \alpha_0)]' [\xi_u - \zeta_A(\alpha - \alpha_0)]}{\sigma_{uu} - 2\Sigma_{uV_A}(\alpha - \alpha_0) + (\alpha - \alpha_0)'\Sigma_{V_A V_A}(\alpha - \alpha_0)}$$

$$\widehat{\alpha}^{LIML} \ \Rightarrow \ \alpha_*^{LIML} = \arg\min_{\alpha \in \mathcal{A}} S_*^{LIML}(\alpha)$$

2.4 Weak instrument robust inference

By Proposition 2, the concentrated objective functions are free of instability effects. Thus they can be used to construct the weak-instrument robust inference for α.

Proposition 3: *Consider the regression model described in (5) and (6). Under Assumptions 2.A to 2.D,* $S_T^{LIML}(\alpha_0, \widehat{\beta}(\alpha_0)) \Rightarrow \chi^2_{p-n_B}$*, where* S_T^{LIML} *is defined in (17),* p *and* n_B *are dimensions of* Z_t *and* Y_{Bt} *respectively.*

Comments:

1. Inference for α

The asymptotic χ^2 distribution of $S_T^{LIML}(\alpha_0, \widehat{\beta}(\alpha_0))$ is a direct consequence of Proposition 2(b). According to Proposition 3, to test for the hypothesis $\alpha = \alpha_0$ (where α_0 is the hypothetical average value of the weakly identified parameter path α_t), we reject if $S_T^{LIML}(\alpha_0, \widehat{\beta}(\alpha_0))$ exceeds the appropriate χ^2 critical value. A confidence set for α alone can be constructed by inverting $S_T^{LIML}(\alpha_0, \widehat{\beta}(\alpha_0))$. Proposition 3 is essentially Stock and Wright (2000 Theorem 3) applied to the linear model.

Note that for the extreme case where all parameters are weakly identified, β does not exist and $n_B = 0$. So we will have $S_T^{LIML}(\alpha_0) \Rightarrow \chi^2_p$.

2. Inference for β

But it is important to emphasize that the construction of asymptotic test and confidence set for β (which is strongly identified) is difficult, *as long as β is unstable.* To see this,

$$S_T^{LIML}(\alpha_0, \beta_0) \Rightarrow [\xi_u + \lambda_B \int W_\beta(s)ds]'[\xi_u + \lambda_B \int W_\beta(s)ds] \;/\sigma_{uu} \;=\; \chi^2_p + \mathcal{G}(W_\beta) \qquad (18)$$

where $\xi_u'\xi_u/\sigma_{uu}$ is independent of β-instability and distributed χ^2_p. But $\mathcal{G}(W_\beta) = (2\xi_u'\lambda_B \int W_\beta(s)ds + \int W_\beta(s)'ds\lambda_B'\lambda_B \int W_\beta(s)ds)/\sigma_{uu}$ is a random function of β-instability. Thus $S_T^{LIML}(\alpha_0, \beta_0)$ has a nonstandard limiting distribution. This result is very different from that of a stable coefficient model where $S_T^{LIML}(\alpha_0, \beta_0) \Rightarrow \chi^2_p$ (see Stock and Wright 2000 Theorem 2).[7]

[7] If α is unstable but β is stable, then $S_T^{LIML}(\alpha_0, \beta_0) \Rightarrow \chi^2_p$ still holds, although in practice it is difficult to find such applications.

3. Other weak instrument robust statistics

There are other statistics, including Anderson and Rubin's (1949) statistic (denoted by AR_T) and Kleibergen's (2005) statistic (denoted by K_T), which share the same spirit as $S_T^{LIML}(\alpha_0, \widehat{\beta}(\alpha_0))$ and can be used to construct weak identification robust inference. In the context of the linear IV model of (5) and (6),

$$AR_T(\alpha, \widehat{\beta}(\alpha)) = \frac{(y - Y_A\alpha - Y_B\widehat{\beta}(\alpha))' P_z (y - Y_A\alpha - Y_B\widehat{\beta}(\alpha))}{(y - Y_A\alpha - Y_B\widehat{\beta}(\alpha))' M_z (y - Y_A\alpha - Y_B\widehat{\beta}(\alpha)) \ /(T-p)}$$

$$K_T(\alpha, \widehat{\beta}(\alpha)) = \frac{(y - Y_A\alpha - Y_B\widehat{\beta}(\alpha))' P_{\widehat{Y}} (y - Y_A\alpha - Y_B\widehat{\beta}(\alpha))}{(y - Y_A\alpha - Y_B\widehat{\beta}(\alpha))' M_{\widehat{Y}} (y - Y_A\alpha - Y_B\widehat{\beta}(\alpha)) \ /(T-p)}$$

where $P_{\widehat{Y}} = \widehat{Y}(\widehat{Y}'\widehat{Y})^{-1}\widehat{Y}'$ and $M_{\widehat{Y}} = I - P_{\widehat{Y}}$ are modified prediction and residual matrices with $\widehat{Y} = P_z \left[Y - (y - Y_A\alpha - Y_B\widehat{\beta}(\alpha)) \frac{(y - Y_A\alpha - Y_B\widehat{\beta}(\alpha))' M_z}{(y - Y_A\alpha - Y_B\widehat{\beta}(\alpha))' M_z (y - Y_A\alpha) - Y_B\widehat{\beta}(\alpha)} Y \right]$.[8]

Using an argument very similar to that of Proposition 3, under the null hypothesis of $\alpha = \alpha_0$, $AR_T(\alpha_0, \widehat{\beta}(\alpha_0)) \Rightarrow \chi^2_{p-n_B}$ and $K_T(\alpha_0, \widehat{\beta}(\alpha_0)) \Rightarrow \chi^2_{n_A}$ where p, n_A and n_B are dimensions of Z_t, Y_{At} and Y_{Bt} respectively.

3 Nonlinear GMM Models

Results in Section 2 are not limited to linear regressions and can be generalized to a broad class of models and estimators. In this section, we consider the widely studied GMM estimation.

3.1 The model and assumptions

To make our results comparable to those in the existing literature, we use as much as possible the notations from Stock and Wright (2000). Assumptions on parameter

[8] AR_T and K_T are closely related to the S_T^{LIML} statistic. S_T^{LIML} and AR_T differ only in their denominators. Moreover, $S_T^{LIML} = T\left[1 + (T-p)AR_T^{-1}\right]^{-1}$. Kleibergen's K-test is an improved version of the AR-test where (i) the first-stage prediction matrix P_z in AR_T is replaced by a modified prediction matrix $P_{\widehat{Y}}$, (ii) the residual matrix M_z in AR_T is replaced by the corresponding modified residual matrix, and (iii) $\widehat{Y}$ in $P_{\widehat{Y}}$ and $M_{\widehat{Y}}$ is essentially a twisted version of the first-stage prediction of the endogenous variable Y. These modifications improve the power property of the AR-test by reducing the degree of freedom.

instability is the same as those in Section 2.1 for the linear IV model.

Consider a GMM model with p moment conditions $E[\phi(Y_t, \theta_t)] = 0$ where Y_t denotes the data and θ_t is an $n \times 1$ vector of possibly unstable parameters. Moreover, θ_t is partitioned as $\theta_t = (\alpha_t' \ \beta_t')'$ where, as will become clear in our latter discussion, α_t is associated with the weakly identified part of the model and β_t is associated with the strongly identified part of the model. To streamline notation, let $\phi_t(\cdot) = \phi(Y_t, \cdot)$. It will at times be convenient to write functions of θ interchangeable as functions of α and β. The population moment condition is then

$$E[\phi_t(\alpha_t, \beta_t)] = 0. \tag{19}$$

Suppose that in practice, practitioners ignore the parameter instability and estimate the model using conventional method for constant parameters, then the estimator $\widehat{\theta}$ is obtained from the following minimization problem

$$\widehat{\theta} = \arg\min_{\theta\in\Theta} S_T(\theta) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]' M_T(\bar{\theta}_T(\theta)) \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right] \tag{20}$$

where $M_T(\bar{\theta}_T(\theta))$ is an $O_p(1)$ positive definite $p \times p$ weighting matrix, and Θ is the parameter space[9].

Different choices of the weighting matrix lead to different estimators. (i) For the one-step estimator, M_T does not depend on data and parameters. For instance $M_T = I$ where I is an p-dimensional identity matrix; (ii) For the efficient two-step estimation, M_T is based on a preliminary estimator of θ. Hence it does not depend on the θ used for the moment function. (iii) For the efficient continuous updating estimator, the weighting matrix is evaluated at the θ value used for the moment function, so $M_T = V_T^{CUE}(\theta)^{-1}$ where $V_T^{CUE}(\theta) = T^{-1}\sum\left[\phi_t(\theta) - \bar{\phi}(\theta)\right]\left[\phi_t(\theta) - \bar{\phi}(\theta)\right]'$ with $\bar{\phi}(\theta) = T^{-1}\sum\phi_t(\theta)$, and the objective function is

$$S_T^{CUE}(\theta) = \left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right]' V_T^{CUE}(\theta)^{-1}\left[T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta)\right] \tag{21}$$

To facilitates our asymptotic analysis, we make the following assumptions. α_t and β_t in (19) obey Assumption 2.B of Section 2.1. The characterization of weak and strong identifications in the nonlinear model is given below.

[9]The notation of $M_T(\bar{\theta}_T(\theta))$ is used to characterize the possible dependence of the weighting matrix on parameters in the moment function.

Assumption 3.A: *The moment function $\phi_t(\theta)$ is assumed to be continuously differentiable and satisfies:*

(a) $E\left[\frac{\partial \phi_t(\theta)}{\partial \alpha'}\right] = T^{-1/2} m_{\alpha,t}(\theta)$ *and*

(b) $E\left[\frac{\partial \phi_t(\theta)}{\partial \beta'}\right] = m_\beta(\theta)$

where $m_{\alpha,t}(\theta)$ and $m_\beta(\theta)$ are $p \times n_A$ and $p \times n_B$ matrices satisfying

(c) $\sup_{t,\theta\in\Theta} ||m_{\alpha,t}(\theta)|| < \infty$ *and* $T^{-1}\sum_{t=1}^{T} m_{\alpha,t}(\theta) \to m_\alpha(\theta)$

uniformly in $\theta \in \Theta$, where $m_\alpha(\theta)$ is a matrix of bounded function; and

(d) $m_\beta(\theta)$ is a continuous $p \times n_B$ matrix with full column rank.

From part (b) of Assumption 3.A, the notation that β is strongly identified is characterized by assuming that the derivative of the moment condition in β does not depend on sample size T. This follows the conventional identification condition in the standard GMM literature. On the other hand, Part (a) of Assumption 3.A adopts the device of linking the expectation of the derivative of the moment condition with respect to the weakly identified parameters, α, to the sample size. This entails considering a sequence of models in which $E[\phi_t(\theta)]$ depends on T. Parts (a) and (b) are a generalization of Assumption 2.C in Section 2.1 for the linear models[10].

Assumption 3.B: $T^{-1/2}\sum_{t=1}^{T}\phi_t(\theta_t) \Rightarrow \Psi_0$ *holds where Ψ_0 is a Gaussian process $\Psi_0 = \mathcal{N}(0, \Omega_\phi)$ with $\Omega_\phi = E[\phi(\theta_t)\phi(\theta_t)']$.*

This assumption states a multivariate Central Limited Theorem to hold for the scaled sample average of the moment function, evaluated at the true time-varying parameters. Given the GMM population moment condition (19), this is a natural condition[11].

[10]To see this, by setting $\phi_t(\theta) = Z_t(y_t - Y'_{At}\alpha - Y'_{Bt}\beta)$ (which is the moment function in the linear model where Z_t is the vector of instruments) and applying the reduced form in (6), we get $E\left[\frac{\partial \phi_t(\theta)}{\partial \alpha'}\right] = -Q_{zz}C_A/\sqrt{T}$ and $E\left[\frac{\partial \phi_t(\theta)}{\partial \beta'}\right] = -Q_{zz}\Pi_B$. Note that for linear models, the expected derivatives of the moment function are independent of the structural parameters (namely α and β). But for nonlinear models, the expected derivatives are allowed to depend on the structural parameters.

[11]For instance, in the linear GMM model of Section 2.1, $\phi_t(\theta_t) = Z_t u_t$ where Z_t is the vector of instruments and u_t is the error term in the structural regression model. Then Assumption 3.B reduces to $T^{-1/2}\sum Z_t u_t \Rightarrow \mathcal{N}(0, Q_{zz}\sigma_{uu})$.

The next assumption is that the scaled sample averages of partial derivatives of the moment condition obey some Central Limit Theorem. Define $\psi_t(\theta) = [\psi_{\alpha,t}(\theta)\ \psi_{\beta,t}(\theta)]$, where $\psi_{\alpha,t}(\theta) = \frac{\partial\phi_t(\theta)}{\partial\alpha'} - E\left[\frac{\partial\phi_t(\theta)}{\partial\alpha'}\right]$ and $\psi_{\beta,t}(\theta) = \frac{\partial\phi_t(\theta)}{\partial\beta'} - E\left[\frac{\partial\phi_t(\theta)}{\partial\beta'}\right]$. So $\psi_t(\theta)$ is the re-centered derivative of the moment function.

Assumption 3.C: $\psi_t(\theta)$ *satisfies the following requirement:*

(a) $T^{-1/2}\sum_{t=1}^{T}\psi_{\alpha,t}(\theta) \Rightarrow \Psi_\alpha(\theta)$

(b) $T^{-1/2}\sum_{t=1}^{T}\psi_{\beta,t}(\theta) \Rightarrow \Psi_\beta(\theta)$

where $\Psi_\alpha(\theta)$ *and* $\Psi_\beta(\theta)$ *are mean-zero Gaussian stochastic processes on* Θ*.*

Since identification also depends on the weighting matrix M_T, we assume that it has a positive definite uniform limit.

Assumption 3.D: M_T *is positive definite and* $M_T(\theta) \to M(\theta)$ *uniformly in* θ *where* $M(\theta)$ *is a symmetric nonrandom* $p \times p$ *matrix that is continuous in* θ *and is positive definite for all* $\theta \in \Theta$*.*

Finally, we impose some continuity requirements.

Assumption 3.E: *For any sequence of positive numbers* $\delta_T \to 0$*,*

(a) $T^{-1}\sum_{t=1}^{T}\sup_{||\theta-\theta_0||<\delta_T}||\psi_t(\theta) - \psi_t(\theta_0)|| \to 0$

(b) $T^{-1}\sum_{t=1}^{T}\sup_{||\theta-\theta_0||<\delta_T}||m_{\alpha,t}(\theta) - m_{\alpha,t}(\theta_0)|| \to 0$

(c) $\sup_{||\theta-\theta_0||<\delta_T}||m_\beta(\theta) - m_\beta(\theta_0)|| \to 0$

Assumption 3.E is an assumption of uniform continuity of ψ_t, $m_{\alpha,t}$ and m_β at θ_0. Take $\psi_t(\theta)$ as an example, the assumption holds if, say, for $i = 1, ..., p$ and $j = 1, ..., n$,

$$T^{-1}\sum_{t=1}^{T}\sup_{||\theta-\theta_0||<\delta_T}\left[E\left(\psi_{ijt}(\theta) - \psi_{ijt}(\theta_0)\right)^2\right]^{1/2} \to 0$$

where $\psi_{ijt}(\theta) = \frac{\partial\phi_{it}(\theta)}{\partial\theta_j} - E\frac{\partial\phi_{it}(\theta)}{\partial\theta_j}$. We could easily replace Assumption 3.E by sufficient conditions that ensures the smoothness of $\psi_t(\theta)$. For example, if we assume that $\phi_t(\theta)$ is differentiable to the second order, and $\sup_{||\theta-\theta_0||<\delta_T} E\left|\frac{\partial^2\phi_{it}(\theta)}{\partial\theta_{j_1}\theta_{j_2}}\right|^2 < \infty$ for $i = 1, ..., p$ and $j_1, j_2 = 1, ..., n$, then Assumption 3.E is satisfied.

It is worthwhile to point out that one implication of the above assumptions, especially Assumptions 3.A and 3.E, is that

$$T^{-1}\sum_{t=1}^{T} E(\phi_t(\theta)) = m_{1T}(\theta)/\sqrt{T} + m_2(\beta) + o(1) \tag{22}$$

where $m_{1T}(\theta) = m_\alpha(\bar{\alpha}, \beta_0)(\alpha-\alpha_0)$ and $m_2(\beta) = m_\beta(\theta_0)(\beta-\beta_0)$ where $\bar{\alpha}$ is an intermediate point between α_0 and α. (22) corresponds to Assumption C in Stock and Wright (2000), and it is derived in the appendix. In other words, (22) justifies Stock-Wright (2000) Assumption C in an unstable coefficient environment.

One reason why we do not use (22) as a high-level assumption is that the addition of parameter instability to the GMM model and its interaction with weak and strong identifications complicates the derivation of the asymptotic results. Using the lower-level conditions as we do here circumvents this difficulty. For instance, Assumptions 3.A, 3.C and 3.E are mainly used to obtain limits of the nonlinear interaction terms of instabilities and weak and strong identifications. These interaction terms are absent in either stable GMM models with weak identification[12]; or unstable GMM models with strong identification[13]; or the unstable linear model with weak identification studied in Section 2 of this paper.

3.2 Asymptotic results

We adopt the following additional notation:

$$\begin{aligned}
\widehat{\beta}(\alpha) &= \arg\min_{\beta\in B} S_T(\alpha, \beta) \qquad (23)\\
\Psi(\alpha, \beta_0) &= \Psi_0 + \Psi_\alpha(\bar{\alpha}, \beta_0)(\alpha - \alpha_0)\\
m_1(\alpha, \beta) &= m_\alpha(\bar{\alpha}, \beta_0)(\alpha - \alpha_0)\\
R(\beta_0) &= m_\beta(\theta_0)
\end{aligned}$$

[12]For instance, see Staiger and Stock (1997) and Stock and Wright (2000).

[13]For instance, see Ghysels and Hall (1990a,b), Li (2008) and Li and Müller (2008).

where $\bar{\alpha}$ is an intermediate point between α_0 and α; Ψ_0 is defined in Assumption 3.B, $\Psi_\alpha(\cdot,\cdot)$ is defined in Assumption 3.C; $m_\alpha(\cdot,\cdot)$ and $m_\beta(\cdot)$ are defined in Assumption 3.A. Let $\beta = \beta_0 + b/\sqrt{T}$.

Proposition 4: *Consider the GMM model described in (19) and (20). Under Assumptions 2.B and 3.A − 3.E, the following limits hold jointly.*

(a) Scaled sample moment condition at $(\alpha, \beta_0 + b/\sqrt{T})$*:*

$$T^{-1/2}\sum_{t=1}^{T}\phi_t(\alpha,\beta_0+b/\sqrt{T}) \Rightarrow \Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R(\beta_0)b - R(\beta_0)\int_0^1 W_\beta(s)ds$$

(b) GMM objective function at $(\alpha, \beta_0 + b/\sqrt{T})$*:*

$$S_T(\alpha,\beta_0+b/\sqrt{T}) \Rightarrow \left[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R(\beta_0)b - R(\beta_0)\int_0^1 W_\beta(s)ds\right]'$$
$$\times M\left[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R(\beta_0)b - R(\beta_0)\int_0^1 W_\beta(s)ds\right]$$

where M *is the limit of the weighting matrix* M_T*.*

(c) Scaled concentrated sample moment condition at $(\alpha, \widehat{\beta}(\alpha))$*:*

$$T^{-1/2}\sum_{t=1}^{T}\phi_t(\alpha,\widehat{\beta}(\alpha)) \Rightarrow \left[I - R(\beta_0)\left(R(\beta_0)'MR(\beta_0)\right)^{-1}R(\beta_0)'M\right]\left[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)\right]$$

(d) Concentrated GMM objective function at $(\alpha, \widehat{\beta}(\alpha))$*:*

$$S_T(\alpha,\widehat{\beta}(\alpha)) \Rightarrow S^*(\alpha) = \left[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)\right]' W(\alpha,\beta_0)\left[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)\right]$$

where $W(\alpha,\beta_0) = M - MR(\beta_0)\left[R(\beta_0)'MR(\beta_0)\right]^{-1}R(\beta_0)'M$

(e) GMM estimator: if $S^*(\alpha)$ *in (d) has a unique minimum, then*

$(\widehat{\alpha}, T^{1/2}(\widehat{\beta}-\beta_0)) \Rightarrow (a^*, b^*)$ *where*

$\alpha^* = \arg\min_{\alpha\in A} S^*(\alpha)$

$b^*(\alpha^*) = -\left[R(\beta_0)'MR(\beta_0)\right]^{-1}R(\beta_0)'M\left[\Psi(\alpha^*,\beta_0)+m_1(\alpha^*,\beta_0)\right] + \int_0^1 W_\beta(s)ds.$

Parts (a) and (b) of Proposition 4 provide the limits of the scaled sample moment condition and objective function at arbitrary α and β. Parts (c) and (d) provides the limits of the sample moment condition and objective function after the strongly identified parameter vector, β, is concentrated out.

As in the linear model of Section 2.1, with the presence of parameter instability, the distribution of $\widehat{\alpha}$ is only affected by weak identification, but not by any parameter instability. However, the asymptotic distribution of $\widehat{\beta}$ is influenced by both weak identification effect spilled over from the α component and the instability effect from the β component itself. Note that α-instability has no effect on the estimation of β.

Finally, Proposition 4 can also be used to obtain explicit expressions for some commonly-used specific GMM estimators. For instance, the limiting objective function for continuous updating GMM can be obtained by replacing the general limiting weighting matrix M in Proposition 4(d) by the limit of $\left[V_T^{CUE}(\alpha, \widehat{\beta}(\alpha))\right]^{-1}$ which gives

$$\begin{aligned} S_T^{CUE}(\alpha, \widehat{\beta}(\alpha)) \;&\Rightarrow\; \left[\widetilde{\Psi}(\alpha, \beta_0) + \widetilde{m}_1(\alpha, \beta_0)\right]' \times \left[I - \widetilde{R}(\alpha, \beta_0)(\widetilde{R}(\alpha, \beta_0)'\widetilde{R}(\alpha, \beta_0))^{-1}\widetilde{R}(\alpha, \beta_0)'\right] \\ &\quad \times \left[\widetilde{\Psi}(\alpha, \beta_0) + \widetilde{m}_1(\alpha, \beta_0)\right] \end{aligned} \tag{24}$$

where $\widetilde{\Psi}(\alpha, \beta_0) = [V_*^{CUE}]^{-1/2}\Psi(\alpha, \beta_0)$, $\widetilde{m}_1(\alpha, \beta_0) = [V_*^{CUE}]^{-1/2}m_1(\alpha, \beta_0)$ and $\widetilde{R}(\alpha, \beta_0) = [V_*^{CUE}]^{-1/2}R(\beta_0)$ where $V_*^{CUE} = \lim_{T\to\infty} V_T^{CUE}(\alpha, \widehat{\beta}(\alpha))$.

In general the distribution of $S_T^{CUE}(\alpha, \widehat{\beta}(\alpha))$ is nonstandard due to the presence of weak identification. However, it can be shown that when (24) is evaluated at α_0, it leads to a central χ^2 distribution.

Proposition 5: *Consider the continuous updating GMM model described in (21). Under Assumptions 2.B and 3.A − 3.E,* $S_T^{CUE}(\alpha_0, \widehat{\beta}(\alpha_0)) \Rightarrow \chi^2_{p-n_B}$ *where* p *is the number of the moment restrictions and* n_B *is the number of the strongly identified parameters.*

Several comments are in order. First, even in the presence of parameter instability, valid inference of α_0 can be made by concentrating out β.

Second, unlike the GMM model with stable coefficients, $S_T^{CUE}(\alpha_0, \beta_0)$ does not have a central χ^2 distribution,

$$\begin{aligned} S_T^{CUE}(\alpha_0, \beta_0) \;&\Rightarrow\; \left[\Psi_0 - R(\beta_0)\int W_\beta(s)ds\right]' V^{CUE}(\alpha_0, \beta_0)^{-1}\left[\Psi_0 - R(\beta_0)\int W_\beta(s)ds\right] \\ &= \chi^2_p + \mathcal{G}(W_\beta) \end{aligned} \tag{25}$$

where $\mathcal{G}(W_\beta) = -2\Psi_0' R(\beta_0)\int W_\beta(s)ds + \int W_\beta(s)'ds R(\beta_0)'V(\alpha_0, \beta_0)^{-1}R(\beta_0)\int W_\beta(s)ds$.[14]

[14](25) is the nonlinear counterpart of (18) of Section 2.3.

Therefore, in the presence of parameter instability in β, we can only make valid inference on the weakly identified parameter α, but not the strongly identified parameter β.

4 Numerical Results

This section examines the accuracy of our asymptotic results of the previous sections in finite samples. We do so by conducting two Monte Carlo experiments. The first experiment concerns the scalar time series IV model discussed in Section 2.2. The second experiment concerns GMM inference in a prototype model with rational expectations, similar to the examples in introduction.

4.1 A scalar IV regression

The data, including the dependent variable y_t, the regressor Y_t and the instrument Z_t, are generated by the following model:

$$\begin{aligned} y_t &= \alpha_t Y_t + u_t && \text{with } \ u_t = (\varepsilon_{1t} + \varepsilon_{2t})/2 \\ Y_t &= (1/\sqrt{T}) Z_t + V_t && \text{with } \ V_t = (\varepsilon_{1t} + \varepsilon_{3t})/2 \\ Z_t &= \varepsilon_{4t} \\ \alpha_t &= 1 + \lambda/\sqrt{T}\, W(t/T) \end{aligned}$$

where $(\varepsilon_{1t}\ \varepsilon_{2t}\ \varepsilon_{3t}\ \varepsilon_{4t})' \sim i.i.d.\ \mathcal{N}(0, I_4)$ and $W(\cdot)$ is a standard Brownian motion that is uncorrelated with the ε's.

In this setup, Y_t is correlated with u_t via ε_{1t}. Instability in α_t follows a random walk process. λ controls the size of the instability: $\lambda = 0$ indicates no instability; a small value of λ ($\lambda < 8$) corresponds to moderate amount of instability; a large λ ($\lambda > 8$) corresponds to large amount of instability[15]. In addition, $T^{-1} \sum \alpha_t \xrightarrow{p} 1$, that is, the average value of α_t path is 1. Finally, $\text{cov}(Y_t, Z_t) = T^{-1/2}$ so that it is a small number for a given sample size T, and hence weak instrument issue arises.

[15] For sample size typically encountered in empirical work, the values of λ smaller than 8 are regarded as more empirically relevant than the larger λ values. See Stock and Watson (1996) for more evidence.

The focus of this Monte Carlo is to assess the impact of the instability in α_t on the inference of α. The heteroskedasticity robust t statistic testing for the hypothesis $\alpha_0 = 1$ is computed for this purpose. The asymptotic theory predicts that (i) the distribution of the t statistic will not converge to $\mathcal{N}(0,1)$ in the presence of weak instrument, and there will be size distortion; (ii) the distribution of the t statistic will be unaffected by the size of the instability.

The Monte Carlo results are displayed in Figure 1. Panel a of Figure 1 plots the finite sample distribution of the t test when $\lambda = 0$ (no instability), $\lambda = 8$ (moderate instability) and $\lambda = 16$ (very large instability). The sample size is $T = 100$. First, note that $\mathcal{N}(0,1)$ provides a poor approximation to finite sample distribution for all sizes of instability. Second, the finite sample distribution does not change much as instability increases. This is true even for very large instability.

Panel b of Figure 1 displays the relationship between the finite sample size of the t test and the magnitude of the instability when the instrument is weak. First, the t test is oversized for all sizes of instability. Second, size distortion is slightly more severe for larger sample size, perhaps because the instrument becomes weaker as sample size increases. Third, as instability becomes very large, size distortion becomes slightly smaller: as λ increases from 0 to 16, empirical rejection rate of the t test drops from 0.25 to 0.21. One possible explanation is that the interaction term between weak instrument and parameter instability works to offset the pure weak instrument effect in a finite sample, although this offsetting effect is small quantitatively.

Overall, the small sample results are consistent with the theoretical predictions.

4.2 A model with rational expectations

We consider a prototype two-equation model with expected future endogenous variables. It is a simplified version of Example 2 in the introduction.

$$y_t = \omega E_t y_{t+1} + \kappa x_t + \varepsilon_t \tag{26}$$

$$x_t = \rho_{1t} x_{t-1} + \rho_{2t} x_{t-1} + \xi_t \tag{27}$$

where the decision variable, y_t, follows an Euler equation. E_t denotes rational expectations conditional on data t information. The driving variable, x_t, follows an AR(2) process[16]. Note that x_t in the Euler is endogenous as long as ξ_t is correlated with ε_t.

In addition to tractability, economic theory has direct implications for the stability of the parameters: the reduced-form coefficients of (27) are unstable because they are functions of time-varying economic policies. Euler equation (26) is assumed to derive from economic agents' optimization problem. As long as preferences remain constant over time, ω and κ are stable.

Under the condition $\omega < 1$, a unique reduced form of the two equation system can be solved as

$$\begin{aligned} y_t &= \alpha_{1t}x_{t-1} + \alpha_{2t}x_{t-2} + \eta_t \\ x_t &= \rho_{1t}x_{t-1} + \rho_{2t}x_{t-2} + \xi_t \end{aligned} \tag{28}$$

where $\eta_t = \varepsilon_t + \gamma_t\xi_t$ and

$$\alpha_{1t} = \frac{\kappa(\rho_{1t} + \omega\rho_{2t})}{1 - \omega\rho_{1t} - \omega^2\rho_{2t}}, \ \alpha_{2t} = \frac{\kappa\rho_{2t}}{1 - \omega\rho_{1t} - \omega^2\rho_{2t}}, \text{ and } \gamma_t = \frac{\kappa}{1 - \omega\rho_{1t} - \omega^2\rho_{2t}}.$$

(28) serves as the data generating process. GMM estimation is proceeded by replacing $E_t y_{t+1}$ in (26) by y_{t+1}, so that estimation of the two equation system is based on

$$\begin{aligned} y_t &= \omega y_{t+1} + \kappa x_t + \widetilde{\varepsilon}_t \\ x_t &= \rho_{1t}x_{t-1} + \rho_{2t}x_{t-1} + \xi_t \end{aligned} \tag{29}$$

where $\widetilde{\varepsilon}_t = \varepsilon_t - \omega(y_{t+1} - E_t y_{t+1})$. It can be shown that $\widetilde{\varepsilon}_t = \varepsilon_t - \omega\varepsilon_{t+1} - \omega\gamma\xi_{t+1}$ holds, using the data generating process (28). This means the error term $\widetilde{\varepsilon}_t$ entering the GMM population moment restriction has an MA(1) structure, and hence suggests the HAC (heteroskedasticity and autocorrelation consistent) estimation of the variance matrix.

The system (29) is just identified[17]. The assumption $E(\widetilde{\varepsilon}_t, \xi_t | x_{t-1}, x_{t-2}) = 0$ leads to the

[16] This is mainly for tractability because it allows us to derive a closed-form solution of the model that can be used to generate the data.

[17] Too see this, lead the first equation in (28) one period ahead and use the second equation in (28) to substitute out x_t, we get the reduced-form equation for y_{t+1} which is $y_{t+1} = (\alpha_1\rho_1 + \alpha_2)x_{t-1} + \alpha_1\rho_2 x_{t-2} + (\varepsilon_t - \omega\eta_{t+1})$. It is then clear that x_{t-1} and x_{t-2} are the only relevant instruments for the two endogenous regressors, y_{t+1} and x_t. Other instruments such as lagged y's are irrelevant. As a result, the two equation system (29) is just identified.

population moment condition $E[\phi_t(\theta_t)] = 0$ where $\theta_t = [\omega\ \kappa\ \rho_{1t}\ \rho_{2t}]'$ and

$$\phi_t(\theta_t) = \begin{bmatrix} \widetilde{\varepsilon}_t \\ \xi_t \end{bmatrix} \otimes \begin{bmatrix} x_{t-1} \\ x_{t-2} \end{bmatrix} = \begin{bmatrix} (y_t - \omega y_{t+1} - \kappa x_t)x_{t-1} \\ (y_t - \omega y_{t+1} - \kappa x_t)x_{t-2} \\ (x_t - \rho_{1t}x_{t-1} - \rho_{2t}x_{t-2})x_{t-1} \\ (x_t - \rho_{1t}x_{t-1} - \rho_{2t}x_{t-2})x_{t-2} \end{bmatrix}$$

The parameter values in the data generating process are as follows: $\omega = 0.5$, $\kappa = 0.3$, $\text{var}(\varepsilon_t) = 1$, $\text{var}(\xi_t) = 2$, $\text{cov}(\varepsilon_t, \xi_t) = -1$. We consider paths of ρ_{1t} and ρ_{2t} that undergo a discrete shift in the middle of the sample,

$$\rho_{1t} = 0.7 + 5K/\sqrt{T}(1_{[t>T/2]} - 0.5), \text{ and } \rho_{2t} = -0.3 - 2K/\sqrt{T}(1_{[t>T/2]} - 0.5)$$

where $1_{[\cdot]}$ is an indicator function. So the size of the break is $5K/\sqrt{T}$ and $2K/\sqrt{T}$ for ρ_1 and ρ_2 respectively, where the value of K controls the magnitude of the parameter instability. We consider $0 \leq K \leq 3$ in the experiment. $T = 100$ throughout the simulations. Note that (i) the average values of ρ_1 and ρ_2 paths are always 0.7 and -0.3, regardless of the size of the break; (ii) from an empirical point of view, given the average values of ρ_1 and ρ_2 and the sample size, $0 < K \leq 2.5$ corresponds to the range of K that gives small to moderate amount of instability while $K > 2.5$ corresponds to fairly large instability.

The specified parameter values have several implications. First, data generated from (28) are stationary throughout for $0 \leq K \leq 3$. Second, the first equation of (29) is weakly identified[18], while the second equation of (29) is strongly identified. Thus when estimating (29) jointly by GMM, asymptotic theory in Sections 2 and 3 makes the following prediction regarding the conventional t tests testing for $\omega = 0.5$ and $\kappa = 0.3$: (i) the t statistics are not distributed $\mathcal{N}(0,1)$ in large samples, (ii) any departure from $\mathcal{N}(0,1)$ is independent of the instabilities in ρ_1 and ρ_2.

Monte Carlo results are reported in Figure 2. Panels a and b plot the distributions of the t tests for ω and κ for three K values: $K = 0$ (no instability), $K = 1$ (small instability) and $K = 2.5$ (moderate instability). For the purpose of comparison, $\mathcal{N}(0,1)$ is also plotted. Panel c plots the finite sample rejection frequency of the t tests. We observe the following patterns. First, $\mathcal{N}(0,1)$ is a poor approximation to the finite sample

[18] We follow the procedure of Stock and Yogo (2004) and use the minimum-eigenvalue statistic in their paper to test the null hypothesis of weak instruments. In all of our replications, the statistic is below the threshold suggested in Stock and Yogo (2004).

distributions. Particularly, both t statistics are asymmetric. Second, the distributions do not shift much as instability increases. Third, the t tests are severely undersized. When $K = 0$ (no instability), the t tests for ω and κ with a nominal size of 5% reject only 1% and 2% of the time, respectively. Fourth, the empirical rejection rates are very little affected when the instability in ρ_1 and ρ_2 is small or moderate ($0 \leq K \leq 2.5$). But when instability gets very large, say $K = 3$, the two 5% t tests reject 4% of the time, which is pretty close to the nominal size.

In summary, our second Monte Carlo experiment with a weakly identified Euler equation suggests that, for moderate size of parameter instability, asymptotic theory developed in this paper provides very good empirical guidance. But this might not be true if the parameter instability is extremely large in magnitude.

5 Conclusion

This paper addresses the consequences of conventional estimation in an unstable model with weak instruments, and the method to conduct valid inference in such models. We find that under quite general conditions, with the simultaneous presence of weak instrument and parameter instability, weak instrument effect dominates. As a result, inference methods developed in the literature that are robust to weak identification in a stable environment remains useful in situations with structural instability. To be more precise, the standard weak-instrument robust inference methods that ignores the instability remains asymptotically valid for the weakly-identified component of the model, as long as the instability is of moderate magnitude in the sense of not being detectable with probability one.

In practice, it might not always be easy to decide which parameters are stable and which are not. But even in this situation, results presented in this paper considerably broaden the applicability of the existing results dealing with weak identification: When conducting inference on weakly-identified coefficients of interest, it is not necessary to assume that they and other nuisance parameters are constant through time.

6 Appendix

Derivations of (13) – (15)

We derive the limits of the three terms in (13) to (15) one by one. (13) holds according to Assumption 2.D part (c). The limit of the left-hand-side term in (14), which reflects the asymptotic impact of instability in α_t, is

$$\begin{aligned} T^{-1/2}\sum Z_t Y_t(\alpha_t-\alpha_0) &= T^{-1}\sum Z_t Y_t W_{\alpha,T}(t/T) \\ &\Rightarrow E(Z_t Y_t)\int W_\alpha(s)ds \\ &= 0 \end{aligned} \tag{30}$$

where the first equality in (30) follows from Assumption 2.B, and the zero limit follows from (i) Assumption 2.D part (d), and (ii) $E(Z_tY_t)=E(Z_t^2)C/\sqrt{T}+E(Z_tV_t)\to 0$ where $E(Z_tV_t)=0$ follows from Assumption 2.A. The limit of the left-hand-side term in (15) can be derived as

$$\begin{aligned} T^{-1/2}\sum Z_t Y_t &= \left[T^{-1}\sum Z_t^2 C + T^{-1/2}\sum Z_t V_t\right] \\ &\Rightarrow Q_{zz}C+\Psi_{zv} \end{aligned} \tag{31}$$

where the first equality is obtained by using the reduced-form equation in (10) and the limit follows from Assumption 2.D parts (b) and (c).

Proposition 1:

We first establish part (a) of the proposition. The scaled sample moment condition can be decomposed as

$$T^{-1/2}\sum Z_t(y_t-Y_{At}'\alpha-Y_{Bt}'\beta)=A_{1T}+A_{2T}+A_{3T}-A_{4T}-A_{5T}$$

where the five terms are

$$\begin{aligned} A_{1T} &= T^{-1/2}\sum Z_t u_t \\ A_{2T} &= T^{-1/2}\sum Z_t Y_{At}'(\alpha_t-\alpha_0) \\ A_{3T} &= T^{-1/2}\sum Z_t Y_{Bt}'(\beta_t-\beta_0) \\ A_{4T} &= T^{-1/2}\sum Z_t Y_{At}'(\alpha-\alpha_0) \\ A_{5T} &= T^{-1/2}\sum Z_t Y_{Bt}'(\beta-\beta_0) \end{aligned}$$

respectively. A_{1T} has the limit of $T^{-1/2}\sum Z_t u_t \Rightarrow \Psi_{zu}=Q_{zz}^{1/2}\xi_u$ by Assumption 2.D part (c).

A_{2T} captures the impact of instability in α_t on the scaled sample moment condition,

$$\begin{aligned} A_{2T} &= T^{-1}\sum Z_t Y_{At}' W_{\alpha,T}(t/T) \\ &= E(Z_t Y_{At}') \int W_\alpha(s)ds \\ &= 0 \end{aligned}$$

which is derived in identical way as we derive equation (30) in the scalar example. A_{3T} captures the impact of instability in β_t on the scaled sample moment condition,

$$\begin{aligned} A_{3T} &= T^{-1}\sum Z_t Y_{Bt}' W_{\beta,T}(t/T) \\ &\Rightarrow E(Z_t Y_{Bt}') \int W_\beta(s)ds \\ &= Q_{zz}\Pi_B \int W_\beta(s)ds \\ &= Q_{zz}^{1/2}\lambda_B \int W_\beta(s)ds \end{aligned}$$

where the first equality follows from Assumption 2.B; the limit follows from Assumption 2.D part (d); the second equality follows from $E(Z_t Y_{Bt}') = E(Z_t Z_t')\Pi_B + E(Z_t V_{Bt}') = Q_{zz}\Pi_B$, and the last equality uses the definition $\lambda_B = Q_{zz}^{1/2}\Pi_B$. Limit of A_{4T} is

$$\begin{aligned} A_{4T} &= T^{-1/2}\sum Z_t(Z_t'\Pi_A + V_{At})(\alpha - \alpha_0) \\ &= \left[T^{-1}\sum Z_t Z_t' C_A + T^{-1/2}\sum Z_t V_{At}\right](\alpha - \alpha_0) \\ &\Rightarrow [Q_{zz}C_A + \Psi_{zv_A}(\alpha - \alpha_0) \\ &= Q_{zz}^{1/2}(\lambda_A + \xi_{V_A})(\alpha - \alpha_0) \end{aligned}$$

The derivation of A_{4T} above is very similar to that of (31) in the scalar example, and the last equality here uses definitions $\lambda_A = Q_{zz}^{1/2}C_A$ and $\xi_{V_A} \sim N(0, \Sigma_{v_A v_A})$. Limit of A_{5T} is

$$\begin{aligned} A_{5T} &= T^{-1/2}\sum Z_t(Z_t'\Pi_B + V_{Bt})(\beta - \beta_0) \\ &= T^{-1}\sum Z_t(Z_t'\Pi_B + V_{Bt})b \\ &\Rightarrow Q_{zz}\Pi_B b \\ &= Q_{zz}^{1/2}\lambda_B b \end{aligned}$$

where in the last equality we use the definition $\lambda_B = Q_{zz}^{1/2}\Pi_B$.

Combining the limits of A_{1T} to A_{5T}, we get

$$\begin{aligned} & T^{-1/2}\sum Z_t(y_t - Y_{At}'\alpha - Y_{Bt}'\beta) \\ \Rightarrow\ & Q_{zz}^{1/2}\left[\xi_u - (\lambda_A + \xi_{V_A})(\alpha - \alpha_0) - \lambda_B b + \lambda_B \int W_\beta(s)ds\right] \end{aligned} \tag{32}$$

Premultiplying (32) by $Q_{zz}^{-1/2}$ on both sides yields Proposition 1 part (a).

Based on the limiting scaled moment condition in (32), we can derive the limiting objective functions at $(\alpha, \beta_0 + b/\sqrt{T})$. Take the TSLS estimation for example,

$$
\begin{aligned}
S_*^{TSLS}(\alpha, b) &= \lim_{T\to\infty} \left[T^{-1/2}\sum Z_t(y_t - Y_{At}'\alpha - Y_{Bt}'\beta)\right]' \\
&\quad \times \left[T^{-1}\sum Z_tZ_t'\right]^{-1} \left[T^{-1/2}\sum Z_t(y_t - Y_{At}'\alpha - Y_{Bt}'\beta)\right] \\
&\Rightarrow \left[Q_{zz}^{1/2}\left(\xi_u - (\lambda_A + \xi_{V_A})(\alpha - \alpha_0) - \lambda_B b + \lambda_B \int W_\beta(s)ds\right)\right]' \\
&\quad \times Q_{zz}^{-1}\left[Q_{zz}^{1/2}\left(\xi_u - (\lambda_A + \xi_{V_A})(\alpha - \alpha_0) - \lambda_B b + \lambda_B \int W_\beta(s)ds\right)\right] \\
&= \left[\xi_u - (\lambda_A + \xi_{V_A})(\alpha - \alpha_0) - \lambda_B b + \lambda_B \int W_\beta(s)ds\right]' \\
&\quad \times \left[\xi_u - (\lambda_A + \xi_{V_A})(\alpha - \alpha_0) - \lambda_B b + \lambda_B \int W_\beta(s)ds\right] \qquad (33)
\end{aligned}
$$

which gives part (b) of Proposition 1.

The limiting objective function of LIML estimation at $(\alpha, \beta_0 + b/\sqrt{T})$ in part (c) can be derived in a similar matter. But we need to derive the limit of $T^{-1}(y - Y_A\alpha - Y_B\beta)'(y - Y_A\alpha - Y_B\beta)$, which is the denominator of $S_*^{LIML}(\alpha, b)$,

$$
\begin{aligned}
& T^{-1}\sum(y_t - Y_{At}'\alpha - Y_{Bt}'\beta)^2 \\
= \;& T^{-1}\sum\left[u_t - Y_{At}'(\alpha - \alpha_0) + Y_{At}'(\alpha_t - \alpha_0) - Y_{Bt}'(\beta - \beta_0) + Y_{Bt}'(\beta_t - \beta_0)\right]^2 \\
= \;& B_{1T} + ... + B_{5T} - 2(B_{6T} + ... + B_{11T}) + 2(B_{12T} + ... + B_{15T}).
\end{aligned}
$$

Among the fifteen terms, there are three terms which have nonzero limits,

$$
\begin{aligned}
B_{1T} &= T^{-1}\sum u_t^2 \to \sigma_{uu} \\
B_{2T} &= T^{-1}\sum(\alpha - \alpha_0)'Y_{At}Y_{At}'(\alpha - \alpha_0) \\
&\to (\alpha - \alpha_0)'\Sigma_{V_AV_A}(\alpha - \alpha_0) \\
B_{6T} &= T^{-1}\sum u_tY_{At}'(\alpha - \alpha_0) \\
&\to \Sigma_{uV_A}(\alpha - \alpha_0)
\end{aligned}
$$

The other terms all have zero limits. To see this, limits of B_{3T} to B_{5T}

$$\begin{aligned}
B_{3T} &= T^{-1}\sum(\beta-\beta_0)'Y_{Bt}Y_{Bt}'(\beta-\beta_0)\\
&= T^{-1}b'T^{-1}\sum Y_{Bt}Y_{Bt}'b\\
&\Rightarrow \lim_{T\to\infty} T^{-1}b'[\Pi_B'Q_{zz}\Pi_B+\Sigma_{V_BV_B}]b=0\\
B_{4T} &= T^{-1}\sum(\alpha_t-\alpha_0)'Y_{At}Y_{At}'(\alpha_t-\alpha_0)\\
&= T^{-2}\sum W_{\alpha,T}(t/T)'Y_{At}Y_{At}'W_{\alpha,T}(t/T)\\
&\Rightarrow \lim_{T\to\infty} T^{-1}\int W_\alpha(s)'\Sigma_{V_AV_A}W_\alpha(s)ds=0\\
B_{5T} &= T^{-1}\sum(\beta_t-\beta_0)'Y_{Bt}Y_{Bt}'(\beta_t-\beta_0)\\
&= T^{-2}\sum W_{\beta,T}(t/T)'Y_{Bt}Y_{Bt}'W_{\beta,T}(t/T)\\
&\Rightarrow \lim_{T\to\infty} T^{-1}\int W_\beta(s)'(\Pi_B'Q_{zz}\Pi_B+\Sigma_{V_BV_B})W_\beta(s)ds=0
\end{aligned}$$

The limits of B_{7T} to B_{11T} are

$$\begin{aligned}
B_{7T} &= T^{-1}\sum u_tY_{Bt}'(\beta-\beta_0)\\
&= \lim_{T\to\infty} T^{-1/2}\Sigma_{uV_B}=0\\
B_{8T} &= T^{-1}\sum(\alpha-\alpha_0)'Y_{At}Y_{At}'(\alpha_t-\alpha_0)\\
&= T^{-1/2}(\alpha-\alpha_0)'T^{-1}\sum Y_{At}Y_{At}'W_{\alpha,T}(t/T)\\
&\Rightarrow \lim_{T\to\infty} T^{-1/2}(\alpha-\alpha_0)'\Sigma_{V_AV_A}\int W_\alpha(s)ds=0\\
B_{9T} &= T^{-1}\sum(\alpha-\alpha_0)'Y_{At}Y_{Bt}'(\beta_t-\beta_0)\\
&= T^{-1/2}(\alpha-\alpha_0)'T^{-1}\sum Y_{At}Y_{Bt}'W_{\beta,T}(t/T)\\
&\Rightarrow \lim_{T\to\infty} T^{-1/2}(\alpha-\alpha_0)'\Sigma_{V_AV_B}\int W_\beta(s)ds=0\\
B_{10T} &= T^{-1}\sum(\beta-\beta_0)'Y_{Bt}Y_{At}'(\alpha_t-\alpha_0)\\
&= T^{-1}b'T^{-1}\sum Y_{Bt}Y_{At}'W_{\alpha,T}(t/T)\\
&\Rightarrow \lim_{T\to\infty} T^{-1}b'\Sigma_{V_AV_B}'\int W_\alpha(s)ds=0\\
B_{11T} &= T^{-1}\sum(\beta-\beta_0)'Y_{Bt}Y_{Bt}'(\beta_t-\beta_0)\\
&= T^{-1}b'T^{-1}\sum Y_{Bt}Y_{Bt}'W_{\beta,T}(t/T)\\
&\Rightarrow \lim_{T\to\infty} T^{-1}b'(\Pi_B'Q_{zz}\Pi_B+\Sigma_{V_BV_B})\int W_\alpha(s)ds=0
\end{aligned}$$

The limits of B_{12T} to B_{15T} are

$$
\begin{aligned}
B_{12T} &= T^{-1}\sum u_t Y_{At}'(\alpha_t - \alpha_0) \\
&= T^{-1/2}T^{-1}\sum u_t Y_{At}' W_{\alpha,T}(t/T) \\
&= \lim_{T\to\infty} T^{-1/2}\Sigma_{uV_A}\int W_\alpha(s)ds = 0 \\
B_{13T} &= T^{-1}\sum u_t Y_{Bt}'(\beta_t - \beta_0) \\
&= T^{-1/2}T^{-1}\sum u_t Y_{Bt}' W_{\beta,T}(t/T) \\
&= \lim_{T\to\infty} T^{-1/2}\Sigma_{uV_B}\int W_\beta(s)ds = 0 \\
B_{14T} &= T^{-1}\sum(\alpha-\alpha_0)' Y_{At} Y_{Bt}'(\beta-\beta_0) \\
&= \lim_{T\to\infty} T^{-1/2}(\alpha-\alpha_0)'[T^{-1/2}C_A' Q_{zz}\Pi_B + \Sigma_{V_A V_B}]b = 0 \\
B_{14T} &= T^{-1}\sum(\alpha_t-\alpha_0)' Y_{At} Y_{Bt}'(\beta_t-\beta_0) \\
&= T^{-2}\sum W_{\alpha,T}(t/T)' Y_{At} Y_{Bt}' W_{\beta,T}(t/T) \\
&\Rightarrow \lim_{T\to\infty} T^{-1}\int W_\alpha(s)\Sigma_{V_A V_B} W_\beta(s)ds = 0
\end{aligned}
$$

Thus, combining these limits, we get

$$T^{-1}\sum(y_t - Y_{At}'\alpha - Y_{Bt}'\beta)^2 \xrightarrow{p} \sigma_{uu} + (\alpha-\alpha_0)'\Sigma_{V_A V_A}(\alpha-\alpha_0) - 2\Sigma_{uV_A}(\alpha-\alpha_0).$$

which is the denominator of the limit of the LIML objective function.

Proposition 2:

For brevity, we only present the derivations for TSLS estimation. The results for LIML estimation can be obtained in very similar way. To concentrate out β, we solve for the optimal b by minimizing the limiting objective function (33). The resulting F.O.C. in b is

$$\lambda_B'\left[\xi_u - \zeta_A(\alpha-\alpha_0) - \lambda_B b + \lambda_B\int W_\beta(s)ds\right] = 0$$

where $\zeta_A = \lambda_A + \xi_{V_A}$ to streamline notation. Rearranging the F.O.C. yields the optimal b as a function of α and the instability in β_t,

$$b^*(\alpha) = (\lambda_B'\lambda_B)^{-1}\lambda_B'\left[\xi_u + \zeta_A(\alpha-\alpha_0)\right] + \int W_\beta(s)ds \tag{34}$$

Substituting (34) into the limiting moment function (32), we get the concentrated limiting moment function,

$$
\begin{aligned}
& \xi_u - \zeta_A(\alpha-\alpha_0) - \lambda_B b^*(\alpha) + \lambda_B \int W_\beta(s)ds \\
= \;& \xi_u - \zeta_A(\alpha-\alpha_0) - \lambda_B \left[(\lambda_B'\lambda_B)^{-1}\lambda_B'\left[\xi_u + \zeta_A(\alpha-\alpha_0)\right] + \int W_\beta(s)ds\right] + \lambda_B \int W_\beta(s)ds \\
= \;& \left[I - \lambda_B(\lambda_B'\lambda_B)^{-1}\lambda_B'\right]\left[\xi_u - \zeta_A(\alpha-\alpha_0)\right] \qquad (35)
\end{aligned}
$$

By defining $M_{\lambda_B} = I - \lambda_B(\lambda_B'\lambda_B)^{-1}\lambda_B'$ we get part (a) of Proposition 3. Plugging (35) into the limiting objective function and noting that matrix M_{λ_B} is idempotent, we get the limit of the TSLS concentrated limiting objective function.

Finally, note that the concentrated limiting objective function is quadratic in α, the limiting distribution of the estimator of α can be computed as the minimizer of the limiting concentrated objective. The corresponding F.O.C. is

$$\zeta_A' M_{\lambda_B}\left[\xi_u - \zeta_A(\alpha_*^{TSLS} - \alpha_0)\right] = 0$$

from which the limiting distribution of the TSLS estimator for α is solved as

$$\alpha_*^{TSLS} - \alpha_0 = (\zeta_A' M_{\lambda_B}\zeta_A)^{-1}\zeta_A' M_{\lambda_B}\xi_u$$

Equation (22)

We start the derivation by the following decomposition,

$$T^{-1}\sum E[\phi_t(\theta)] = A_{1T} + A_{2T}$$

where the expressions of the two terms are

$$
\begin{aligned}
A_{1T} &= T^{-1}\sum E[\phi_t(\theta)-\phi_t(\theta_0)] \\
&= T^{-1}\sum E\left[\phi_t(\alpha,\beta)-\phi_t(\alpha_0,\beta)+\phi_t(\alpha_0,\beta)-\phi_t(\alpha_0,\beta_0)\right] \\
&= B_{1T}+B_{2T} \\
A_{2T} &= T^{-1}\sum E[\phi_t(\theta_t)-\phi_t(\theta_0)] \\
&= T^{-1}\sum E\left[\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\theta'}(\theta_t-\theta_0)\right] \\
&= T^{-1}\sum E\left\{\left[\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\theta'}-E\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\theta'}\right](\theta_t-\theta_0)\right\} \\
&\quad +T^{-1}\sum E\left[\left(E\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\theta'}\right)(\theta_t-\theta_0)\right] \\
&= B_{3T}+B_{4T}+B_{5T}
\end{aligned}
$$

where $\bar{\theta}_t$ is an intermediate point between θ_t and θ_0, and elements B_{1T} to B_{5T} are

$$
\begin{aligned}
B_{1T} &= T^{-1}\sum E\left[\phi_t(\alpha,\beta)-\phi_t(\alpha_0,\beta)\right]=T^{-1}\sum E\left[\frac{\partial\phi_t(\bar{\alpha},\beta)}{\partial\alpha'}\right](\alpha-\alpha_0) \\
B_{2T} &= T^{-1}\sum E\left[\phi_t(\alpha_0,\beta)-\phi_t(\alpha_0,\beta_0)\right]=T^{-1}\sum E\left[\frac{\partial\phi_t(\alpha_0,\bar{\beta})}{\partial\beta'}\right](\beta-\beta_0) \\
B_{3T} &= T^{-1}\sum E\left[\psi_t(\bar{\theta}_t)(\theta_t-\theta_0)\right] \\
B_{4T} &= T^{-1}\sum E\left[\left(E\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\alpha'}\right)(\alpha_t-\alpha_0)\right] \\
B_{5T} &= T^{-1}\sum E\left[\left(E\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\beta'}\right)(\beta_t-\beta_0)\right]
\end{aligned}
$$

where $\bar{\alpha}$ is an intermediate point between α and α_0; $\bar{\beta}$ is an intermediate point between β and β_0. We first study the behavior of A_{1T}. The limits of the two terms in A_{1T} are

$$
\begin{aligned}
B_{1T} &= T^{-1}\sum\left[T^{-1/2}m_{\alpha,t}(\bar{\alpha},\beta)\right](\alpha-\alpha_0) \\
&= T^{-1/2}\sum\left[T^{-1}m_{\alpha,t}(\bar{\alpha},\beta)\right](\alpha-\alpha_0) \\
&= T^{-1/2}m_{\alpha}(\bar{\alpha},\beta)(\alpha-\alpha_0)+o(1) \qquad (36) \\
B_{2T} &= m_{\beta}(\alpha_0,\beta_0)(\beta-\beta_0)+o(1) \qquad (37)
\end{aligned}
$$

Combining (36) and (37), we get

$$
A_{1T}=T^{-1/2}m_{\alpha}(\bar{\alpha},\beta)(\alpha-\alpha_0)+m_{\beta}(\alpha_0,\beta_0)(\beta-\beta_0)+o(1) \qquad (38)
$$

Next we show that the limit of A_{2T} is zero. Note that A_{2T} captures the impact of instability on the quantity of interest $T^{-1}\sum E[\phi_t(\theta)]$. For the first term of A_{2T},

$$\begin{aligned}
B_{3T} &= T^{-1}\sum E\left[\psi_t(\bar{\theta}_t)(\theta_t-\theta_0)\right] \\
&= T^{-1}\sum E\left[\psi_t(\bar{\theta}_t)T^{-1/2}W_{\theta,T}(t/T)\right] \\
&= T^{-1}\sum E\left[\psi_t(\theta_0)T^{-1/2}W_{\theta,T}(t/T)\right] \\
&\quad +T^{-1}\sum E\left[(\psi_t(\bar{\theta}_t)-\psi_t(\theta_0))T^{-1/2}W_{\theta,T}(t/T)\right] \\
&= T^{-1/2}E\left[T^{-1}\sum \psi_t(\theta_0)W_{\theta,T}(t/T)\right]+o(1) \\
&= o(1)
\end{aligned}$$

where (i) the first equality follows from Assumption 2.B for θ_t; (ii) the last equality uses Lemma A.2 of Stock and Watson (1998) which delivers $T^{-1}\sum \psi_t(\theta_0)W_{\theta,T}(t/T)\to E(\psi_t(\theta_0))\int W_\theta(s)ds$, but then note that $E(\psi_t(\theta_0))=0$ according to the definition of function $\psi_t(\cdot)$; (iii) the result that $T^{-1}\sum E[(\psi_t(\bar{\theta}_t)-\psi_t(\theta_0))T^{-1/2}W_{\theta,T}(t/T)]=o(1)$ in the fourth equality is obtained by the following arguments: It is a $p\times 1$ vector with the ith element being

$$\begin{aligned}
& T^{-1}\sum_{t=1}^{T}E\left[(\psi_{it}(\bar{\theta}_t)-\psi_{it}(\theta_0))T^{-1/2}W_{\theta,T}(t/T)\right] \\
= \;& T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n}E\left[(\psi_{ijt}(\bar{\theta}_t)-\psi_{ijt}(\theta_0))T^{-1/2}W_{\theta,jT}(t/T)\right]
\end{aligned}$$

where

$$\begin{aligned}
\psi_{it}(\theta) &= (\psi_{i1,1}(\theta),...,\psi_{in,t}(\theta)) \quad i=1,...,p \\
\psi_{ijt}(\theta) &= \frac{\partial\phi_{it}(\theta)}{\partial\theta_j}-E\left[\frac{\partial\phi_{it}(\theta)}{\partial\theta_j}\right] \quad j=1,...,n
\end{aligned}$$

Thus, for any i, we have

$$\begin{aligned}
& \left|T^{-1}\sum_{t=1}^{T}E\left[(\psi_{it}(\bar{\theta}_t)-\psi_{it}(\theta_0))T^{-1/2}W_{\theta,T}(t/T)\right]\right| \\
\leq \;& T^{-1/2}T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n}\left[E\left(\psi_{ijt}(\bar{\theta}_t)-\psi_{ijt}(\theta_0)\right)^2\right]^{1/2}\left[E(W_{\theta_j,T}(t/T))^2\right]^{1/2} \\
\leq \;& \sup_{0\leq s\leq 1}\left[E(W_{\theta_j}(s))^2\right]^{1/2}T^{-1/2}T^{-1}\sum_{t=1}^{T}\sum_{j=1}^{n}\left[E\left(\psi_{ijt}(\bar{\theta}_t)-\psi_{ijt}(\theta_0)\right)^2\right]^{1/2}
\end{aligned}$$

As a result,

$$
\begin{aligned}
& \left|\left|T^{-1/2}T^{-1}E\left[\sum_{t=1}^{T}(\psi_t(\bar{\theta}_t)-\psi_t(\theta_0))W_{\theta,T}(t/T)\right]\right|\right| \\
\leq \;& T^{-1/2}\sup_{0\leq s\leq 1}||W_\theta(s)||\sup_{\theta}T^{-1}\sum_{t=1}^{T}||\psi_t(\theta)-\psi_t(\theta_0)|| \\
\rightarrow \;& 0
\end{aligned}
$$

by Assumption 3.E part (a). Next, for the second term of A_{2T},

$$
\begin{aligned}
B_{4T} &= T^{-1}\sum E\left[E\left(\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\alpha'}\right)(\alpha_t-\alpha_0)\right] \\
&= T^{-1}\sum E\left[T^{-1/2}m_{\alpha,t}(\bar{\theta}_t)(\alpha_t-\alpha_0)\right] \\
&= T^{-1}\sum E\left[T^{-1/2}m_{\alpha,t}(\bar{\theta}_t)T^{-1/2}W_{\alpha,T}(t/T)\right]
\end{aligned}
$$

where the first equality is based on Assumption 3.A part (a), and the last equality uses Assumption 2.B. Therefore, based on Assumption 3.A part (c),

$$
\begin{aligned}
& ||B_{4T}|| \\
\leq \;& T^{-1}\sum E\left|\left|\left[E\left(E\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\alpha'}\right)(\alpha_t-\alpha_0)\right]\right|\right| \\
\leq \;& T^{-1}\sup_{t,\theta\in\Theta}||m_{\alpha,t}(\theta)||\left|\left|T^{-1}\sum EW_{\alpha,T}(t/T)\right|\right| \\
= \;& O(T^{-1})\rightarrow 0
\end{aligned}
$$

For the third term of A_{2T},

$$
\begin{aligned}
B_{5T} &= T^{-1}\sum E\left[E\left(\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\beta'}\right)(\beta_t-\beta_0)\right] \\
&= T^{-1}\sum E\left[m_\beta(\bar{\theta}_t)(\beta_t-\beta_0)\right] \\
&= m_\beta(\theta_0)T^{-1/2}E\left[T^{-1/2}\sum(\beta_t-\beta_0)\right]+o(1) \\
&= T^{-1/2}m_\beta(\theta_0)E\int W_\beta(s)ds+o(1) \\
&= O(T^{-1/2})\rightarrow 0
\end{aligned}
$$

Thus, we have $A_{2T}=B_{3T}+B_{4T}+B_{5T}\rightarrow 0$ because all three terms of A_{2T} have zero limits. This says, parameter instability has only negligible effect on $T^{-1}\sum E\phi_t(\theta)$. Then, behavior of $T^{-1}\sum E\phi_t(\theta)$ is only determined by A_{1T},

$$
T^{-1}\sum E\phi_t(\theta)=T^{-1/2}m_\alpha(\bar{\alpha},\beta)(\alpha-\alpha_0)+m_\beta(\alpha_0,\beta_0)(\beta-\beta_0)+o(1).
$$

Proposition 4

We start with part (a) which is the key of most results in this proposition. The scaled sample moment condition $T^{-1/2}\sum \phi_t(\theta)$ can be decomposed as

$$T^{-1/2}\sum \phi_t(\theta) = C_{1T} + C_{2T} + C_{3T} + C_{4T} + C_{5T} + C_{6T} + C_{7T}$$

where

$$\begin{aligned}
C_{1T} &= T^{-1/2}\sum \phi_t(\theta_t) \\
C_{2T} &= T^{-1/2}\sum \left[\frac{\partial \phi_t(\bar{\alpha},\beta)}{\partial \alpha'} - E\left(\frac{\partial \phi_t(\bar{\alpha},\beta)}{\partial \alpha'}\right)\right](\alpha-\alpha_0) \\
C_{3T} &= T^{-1/2}\sum \left[\frac{\partial \phi_t(\alpha_0,\bar{\beta})}{\partial \beta'} - E\left(\frac{\partial \phi_t(\alpha_0,\bar{\beta})}{\partial \beta'}\right)\right](\beta-\beta_0) \\
C_{4T} &= -T^{-1/2}\sum \frac{\partial \phi_t(\bar{\alpha}_t,\beta_t)}{\partial \alpha'}(\alpha_t-\alpha_0) \\
C_{5T} &= -T^{-1/2}\sum \frac{\partial \phi_t(\alpha_0,\bar{\beta}_t)}{\partial \beta'}(\beta_t-\beta_0) \\
C_{6T} &= T^{-1/2}\sum \left[E\left(\frac{\partial \phi_t(\bar{\alpha},\beta)}{\partial \alpha'}\right)\right](\alpha-\alpha_0) \\
C_{7T} &= T^{-1/2}\sum \left[E\left(\frac{\partial \phi_t(\alpha_0,\bar{\beta})}{\partial \beta'}\right)\right](\beta-\beta_0)
\end{aligned}$$

where $\bar{\alpha}$, $\bar{\beta}$, $\bar{\alpha}_t$ and $\bar{\beta}_t$ are intermediate points between α and α_0, β and β_0, α_t and α_0 and β_t and β_0, respectively. The above decomposition is obtained by decomposing $\phi_t(\theta)$ as

$$\begin{aligned}
\phi_t(\theta) \quad = \quad & \phi_t(\alpha_t,\beta_t) + [\phi_t(\alpha,\beta) - \phi_t(\alpha_0,\beta)] + [\phi_t(\alpha_0,\beta) - \phi_t(\alpha_0,\beta_0)] \\
& -[\phi_t(\alpha_t,\beta_t) - \phi_t(\alpha_0,\beta_t)] - [\phi_t(\alpha_0,\beta_t) - \phi_t(\alpha_0,\beta_0)]
\end{aligned}$$

and use the mean-value expansions. Among these seven terms of C_{1T} to C_{7T}, note that C_{4T} captures the instability effects of the weakly identified part of the model. C_{5T} is the instability effects of the strongly identified part of the model.

Next we derive one-by-one the limits of the seven components of $T^{-1/2}\sum \phi_t(\theta)$. For the first three components, we have

$$\begin{aligned}
& C_{1T} \Rightarrow \Psi_0 \\
& C_{2T} = T^{-1/2}\sum \psi_{\alpha,t}(\bar{\alpha},\beta)(\alpha-\alpha_0) \Rightarrow \Psi_\alpha(\bar{\alpha},\beta_0)(\alpha-\alpha_0) \\
& C_{3T} = T^{-1/2}\sum \psi_{\beta,t}(\alpha_0,\bar{\beta})(\beta-\beta_0) \Rightarrow \Psi_\beta(\alpha_0,\beta_0)b/\sqrt{T} \to 0
\end{aligned}$$

where limit of C_{1T} follows from Assumption 3.B; limit of C_{2T} uses Assumption 3.C part (a); limit of C_{3T} uses Assumption 3.C part (b) and $\beta = \beta_0 + b/\sqrt{T}$. To get limit of C_{4T}, we further decompose it into

$$
\begin{aligned}
C_{4T} &= -T^{-1/2}\sum\left[\frac{\phi_t(\bar{\alpha}_t,\beta_t)}{\partial\alpha'}-E\left(\frac{\phi_t(\bar{\alpha}_t,\beta_t)}{\partial\alpha'}\right)\right](\alpha_t-\alpha_0)+T^{-1/2}\sum\left[E\left(\frac{\phi_t(\bar{\alpha}_t,\beta_t)}{\partial\alpha'}\right)\right](\alpha_t-\alpha_0)\\
&= -T^{-1/2}\sum\psi_{\alpha,t}(\bar{\alpha}_t,\beta_t)(\alpha_t-\alpha_0)+T^{-1/2}\sum\left[E\left(\frac{\phi_t(\bar{\alpha}_t,\beta_t)}{\partial\alpha'}\right)\right]\\
&= -T^{-1/2}\sum\left[\psi_{\alpha,t}(\bar{\theta}_t)-\psi_{\alpha,t}(\theta_0)\right](\alpha_t-\alpha_0)-T^{-1/2}\sum\psi_{\alpha,t}(\theta_0)(\alpha_t-\alpha_0)\\
&\quad +T^{-1/2}\sum\left[E\left(\frac{\phi_t(\bar{\alpha}_t,\beta_t)}{\partial\alpha'}\right)\right](\alpha_t-\alpha_0)\\
&= -D_{1T}-D_{2T}+D_{3T}
\end{aligned}
$$

where

$$
\begin{aligned}
D_{1T} &= T^{-1/2}\sum\left[\psi_{\alpha,t}(\bar{\theta}_t)-\psi_{\alpha,t}(\theta_0)\right](\alpha_t-\alpha_0)\\
D_{2T} &= T^{-1/2}\sum\psi_{\alpha,t}(\theta_0)(\alpha_t-\alpha_0)\\
D_{3T} &= T^{-1/2}\sum\left[E\left(\frac{\phi_t(\bar{\alpha}_t,\beta_t)}{\partial\alpha'}\right)\right](\alpha_t-\alpha_0)
\end{aligned}
$$

Note that by Assumption 3.E part (a), we have

$$
\begin{aligned}
||D_{1T}|| &= \left|\left|T^{-1}\sum\left[\psi_{\alpha,t}(\bar{\theta}_t)-\psi_{\alpha,t}(\theta_0)\right]W_{\alpha,T}(t/T)\right|\right|\\
&\leq \sup_{0\leq s\leq 1}||W_\alpha(s)||T^{-1}\sum\sup_{\theta}||\psi_{\alpha,t}(\bar{\theta}_t)-\psi_{\alpha,t}(\theta_0)||
\end{aligned}
$$

Thus, $D_{1T}\to 0$. For the term D_{2T},

$$
\begin{aligned}
D_{2T} &= E(\psi_{\alpha,t}(\theta_0))T^{-1/2}\sum(\alpha_t-\alpha_0)+o_p(1)\\
&\to E(\psi_{\alpha,t}(\theta_0))\int W_\alpha(s)ds\\
&= 0
\end{aligned}
$$

where we use (i) Lemma 2 of Stock and Watson (1998) for the first equality and (ii) $E(\psi_{\alpha,t}(\theta_0))=0$ (according to the definition of function $\psi_{\alpha,t}(\cdot)$) for the second equality. For the term D_{3T}, we have

$$
\begin{aligned}
D_{3T} &= T^{-1}\sum\left[T^{-1/2}m_{\alpha,t}(\bar{\theta}_t)\right]W_{\alpha,T}(t/T)\\
&= T^{-1/2}\left\{T^{-1}\sum\left[m_{\alpha,t}(\bar{\theta}_t)-m_{\alpha,t}(\theta_0)\right]W_{\alpha,T}(t/T)+T^{-1}\sum m_{\alpha,t}(\theta_0)W_{\alpha,T}(t/T)\right\}\\
&= T^{-1/2}\left[m_\alpha(\theta_0)\int W_\alpha(s)ds+o(1)\right]\\
&= O(T^{-1/2})
\end{aligned}
$$

where (i) the first equality uses Assumption 3.A part (a) and Assumption 2.B; (ii) $T^{-1}\sum m_{\alpha,t}(\theta_0)W_{\alpha,T}(t/T)\Rightarrow m_\alpha(\theta_0)\int W_\alpha(s)ds$ in the third equality follows from Lemma A.2

of Stock and Watson (1998) and Assumption 3.A part (c); and (iii) in the third equality we also use the following argument

$$\begin{aligned} & \left|\left|T^{-1}\sum\left[m_{\alpha,t}(\bar{\theta}_t)-m_{\alpha,t}(\theta_0)\right]W_{\alpha,T}(t/T)\right|\right| \\ \leq & \sup_{0\leq s\leq 1}||W_\alpha(s)||T^{-1}\sum\sup_{\theta\in\Theta_0}||m_{\alpha,t}(\theta)-m_{\alpha,t}(\theta_0)|| \\ \rightarrow & \ 0 \end{aligned}$$

which uses Assumption 3.A part (c). Therefore, $C_{4T}=-D_{1T}-D_{2T}+D_{3T}\rightarrow 0$. Limit of C_{5T} can be obtained in a very similar way to that of C_{4T}, that is,

$$\begin{aligned} C_{5T} &= -E_{1T}-E_{2T}+E_{3T} \\ E_{1T} &= T^{-1/2}\sum\left[\psi_{\beta,t}(\bar{\theta}_t)-\psi_{\beta,t}(\theta_0)\right](\beta_t-\beta_0)\rightarrow 0 \\ E_{2T} &= T^{-1/2}\sum\psi_{\beta,t}(\theta_0)(\beta_t-\beta_0)\rightarrow 0 \\ E_{3T} &= T^{-1/2}\sum\left[E\left(\frac{\partial\phi_t(\bar{\theta}_t)}{\partial\beta'}\right)\right](\beta_t-\beta_0) \\ &= T^{-1}\sum m_\beta(\bar{\theta}_t)W_{\beta,T}(t/T) \\ &= T^{-1}\sum\left[m_\beta(\bar{\theta}_t)-m_\beta(\theta_0)\right]W_{\beta,T}(t/T)+T^{-1}\sum m_\beta(\theta_0)W_{\beta,T}(t/T) \\ &= m_\beta(\theta_0)\int W_\beta(s)ds+o(1) \end{aligned}$$

which gives the result that $C_{5T}\Rightarrow m_\beta(\theta_0)\int W_\beta(s)ds$. The limits of C_{6T} and C_{7T} are

$$\begin{aligned} C_{6T} &= T^{-1/2}\sum\left[T^{-1/2}m_{\alpha,t}(\bar{\alpha},\beta)\right](\alpha-\alpha_0)\Rightarrow m_\alpha(\bar{\alpha},\beta_0)(\alpha-\alpha_0) \\ C_{7T} &= T^{-1/2}\sum m_\beta(\alpha_0,\beta)b/\sqrt{T}\Rightarrow m_\beta(\theta_0)b \end{aligned}$$

which follows from Assumption 3.A. Collecting the limits of C_{1T} to C_{7T}, we get

$$\begin{aligned} & T^{-1/2}\sum\phi_t(\alpha,\beta_0+b/\sqrt{T}) \\ = & \ \Psi_0+\Psi_\alpha(\bar{\alpha},\beta_0)(\alpha-\alpha_0)+m_\alpha(\bar{\alpha},\beta_0)(\alpha-\alpha_0)+m_\beta(\theta_0)b-m_\beta(\theta_0)\int W_\beta(s)ds \\ = & \ \Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R(\beta_0)b-R(\beta_0)\int W_\beta(s)ds \end{aligned} \tag{39}$$

where the last equality uses the notations introduced in (23). The result in (39) is part (a) of Proposition 4. Then, substituting (39) into the GMM objective function (20),

$$\begin{aligned} \bar{S}(\alpha,b) &= \lim_{T\rightarrow\infty}\left[T^{-1/2}\sum\phi_t(\alpha,\beta)\right]'M_T(\bar{\theta}(\alpha,\beta))\left[T^{-1/2}\sum\phi_t(\alpha,\beta)\right] \\ &\Rightarrow \left[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R(\beta_0)b-R(\beta_0)\int W_\beta(s)ds\right]' \\ &\quad \times M\left[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R(\beta_0)b-R(\beta_0)\int W_\beta(s)ds\right] \end{aligned} \tag{40}$$

which is part (b) of Proposition 4. Next, we concentrate out b. The F.O.C. of (40) in b is

$$0 = R(\beta_0)'M\left(\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0) + R(\beta_0)b - R(\beta_0)\int W_\beta(s)ds\right)$$

From the F.O.C., b can be solved as a function of α,

$$\begin{aligned}
& b^*(\alpha) \\
= & -[R(\beta_0)'MR(\beta_0)]^{-1}R(\beta_0)'M\left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0) - R(\beta_0)\int W_\beta(s)ds\right] \\
= & -[R(\beta_0)'MR(\beta_0)]^{-1}R(\beta_0)'M\left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0)\right] + \int W_\beta(s)ds
\end{aligned} \tag{41}$$

This gives part (e) of Proposition 4. Next, we substitute $b^*(\alpha)$ in (41) back into the limiting objective function (40). Note that

$$\begin{aligned}
& R(\beta_0)b^*(\alpha) \\
= & -R(\beta_0)[R(\beta_0)'MR(\beta_0)]^{-1}R(\beta_0)'M\left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0)\right] + R(\beta_0)\int W_\beta(s)ds
\end{aligned}$$

As a result, the limit of the concentrated sample moment condition is

$$\begin{aligned}
& \Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0) + R(\beta_0)b^*(\alpha) - R(\beta_0)\int W_\beta(s)ds \\
= & \Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0) - R(\beta_0)[R(\beta_0)'MR(\beta_0)]^{-1}R(\beta_0)'M\left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0)\right] \\
& + R(\beta_0)\int W_\beta(s)ds - R(\beta_0)\int W_\beta(s)ds \\
= & \left[I - R(\beta_0)(R(\beta_0)'MR(\beta_0))^{-1}R(\beta_0)'M\right]\left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0)\right]
\end{aligned} \tag{42}$$

from which we can see that the distortion in $b^*(\alpha)$ induced by the β-instability is canceled out. This is part (c) of Proposition 4. Therefore, the limiting objective function (40) evaluated at $(\alpha, b^*(\alpha))$ can be derived as

$$\begin{aligned}
& S^*(\alpha) = \bar{S}(\alpha, b^*(\alpha)) \\
= & \left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0) + R(\beta_0)b^*(\alpha) - R(\beta_0)\int W_\beta(s)ds\right]' \\
& \times M\left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0) + R(\beta_0)b^*(\alpha) - R(\beta_0)\int W_\beta(s)ds\right] \\
= & \left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0)\right]' \\
& \times \left[I - R(\beta_0)(R(\beta_0)'MR(\beta_0))^{-1}R(\beta_0)'M\right]' M \left[I - R(\beta_0)(R(\beta_0)'MR(\beta_0))^{-1}R(\beta_0)'M\right] \\
& \times \left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0)\right] \\
= & \left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0)\right]' W(\alpha,\beta_0)\left[\Psi(\alpha,\beta_0) + m_1(\alpha,\beta_0]\right)
\end{aligned}$$

where $W(\alpha,\beta_0) = M - MR(\beta_0)(R(\beta_0)'MR(\beta_0))^{-1}R(\beta_0)'M$ is an idempotent matrix. This gives part (d) of Proposition 4.

7 References

Andrews, D., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821 − 856.

Andrews, D., 2003. End-of-point instability tests. Econometrica 71, 1661 − 1694.

Andrews, D. and Ploberger W., 1994. Optimal tests when a nuisance parameter is present only under the alternative, Econometrica 62, 1383 − 1414.

Bai, J., Lumsdaine R. and Stock J., 1998. Testing For and Dating Common Breaks in Multivariate Time Series, Review of Economic Studies 65, 395 − 432.

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

Cogley T. and Sargent T., 2005. Drifts and Volatilities: Monetary Policies and Outcomes in the Post WWII U.S., Review of Economic Dynamics 8, 219 - 255.

Dufour J., Khalaf L. and Kichian M., 2006. Inflation dynamics and the New Keynesian Phillips Curve: an identification robust econometric analysis. Journal of Economic Dynamics and Control 30, 1707 − 1727.

Elliott, G. and Müller, U., 2006. Efficient tests for general persistent time variation in regression coefficients, Review of Economic Studies 73, 907 − 940.

Fuhrer, J. and Rudebusch, G., 2004. Estimating the Euler Equation for Output, Journal of Monetary Economics 51. 1133 − 1153.

Ghysels, E. and Hall, A., 1990a. A test for structural stability of Euler conditions parameters estimated via the generalized method of moments estimator, International Economic Review 31, 355 − 364.

Ghysels, E. and Hall,A., 1990b. Are consumption-based intertemporal capital asset pricing models structural, Journal of Econometrics 45, 121 − 139.

Hansen, B., 2000. Testing for structural change in conditional models, Journal of econometrics 97, 93 − 115.

Kleibergen, F. 2005. Testing parameters in GMM without assuming that they are identified. Econometrica 73, 1103 − 1123.

Li, H. 2008. Estimation and testing of Euler equation models with time-varying reduced-form coefficients. Journal of econometrics 142, 425 − 448.

Li, H. and Müller U., 2008. Valid inference in partially unstable GMM models, Manuscript (Brandeis University).

Linde, J. 2001. Testing for the Lucas-critique: a quantitative investigation, American economic review 91, 986 − 1005.

Ma, A., 2002. GMM estimation of the new Phillips curve, Economic letters 76, 411 − 417.

Mavroeidis, S., 2004. Weak identification of forward-looking models in monetary economics. Oxford Bulletin of Economics and Statistics, vol. 66.

Mavroeidis, S., 2005. Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve. Journal of Money Credit and Banking, vol 37(3).

Nyblom, J., 1989. Testing for the constancy of parameters over time, Journal of American Statistical Association 78, 856 − 864.

Sowell, F. 1996. Optimal tests for parameter instability in the generalized method of moments framework. Econometrica 64, 1085 − 1107.

Staiger, D. and Stock, J., 1997. Instrumental variables regression with weak instruments. Econometrica 65, 557 − 586.

Stock, J. and Watson, M., 1996. Evidence on structural instability in macroeconomic time series relations, Journal of Business and Economic Statistics 14, 11 − 29.

Stock, J. and Watson, M., 1998. Median unbiased estimation of coefficient variance in a time-varying parameter model, Journal of American Statistical Association 93, 349 − 358.

Stock, J. and Wright J. 2000. GMM with weak identification. Econometrica 68, 1055 − 1086.

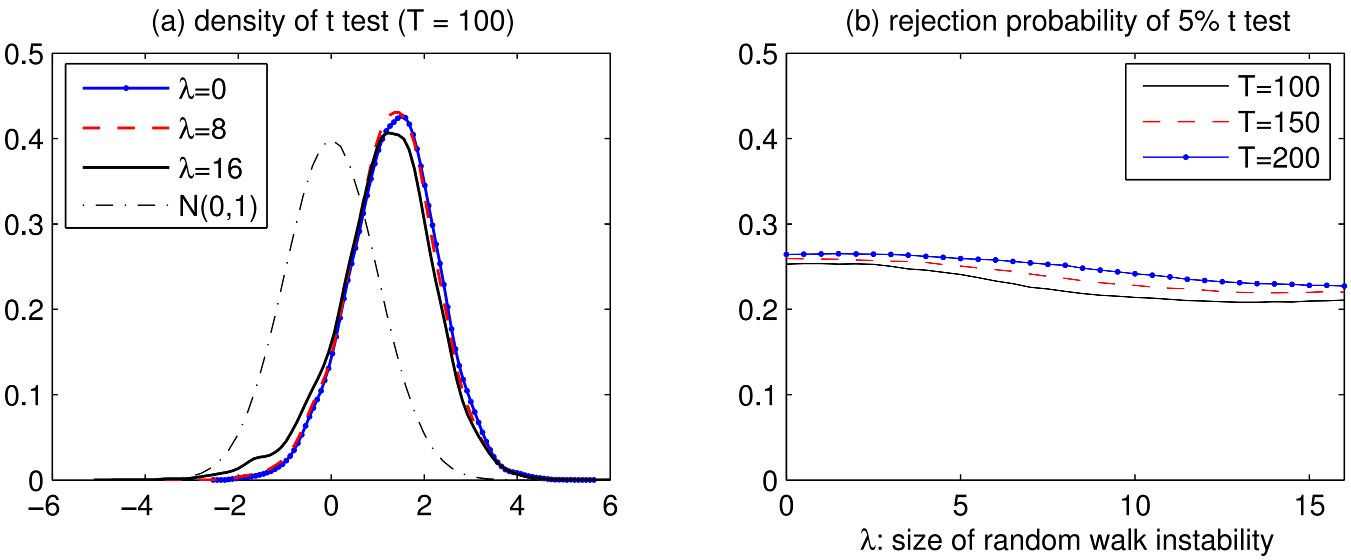
Stock, J., Wright J. and Yogo M. 2002. A survey of weak instruments and weak identification in generalized method of moments. Journal of business and economic statistics 20(4), 518 − 529.

Stock, J. and Yogo M. 2005a. Asymptotic Distributions of Instrumental Variables Statistics with Many Weak Instruments, in D. Andrews and J. Stock, eds., Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, 109 − 120.

Stock, J. and Yogo M. 2005b. Testing for Weak Instruments in Linear IV Regression, in D. Andrews and J. Stock, eds., Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg. Cambridge University Press, 80 − 108.

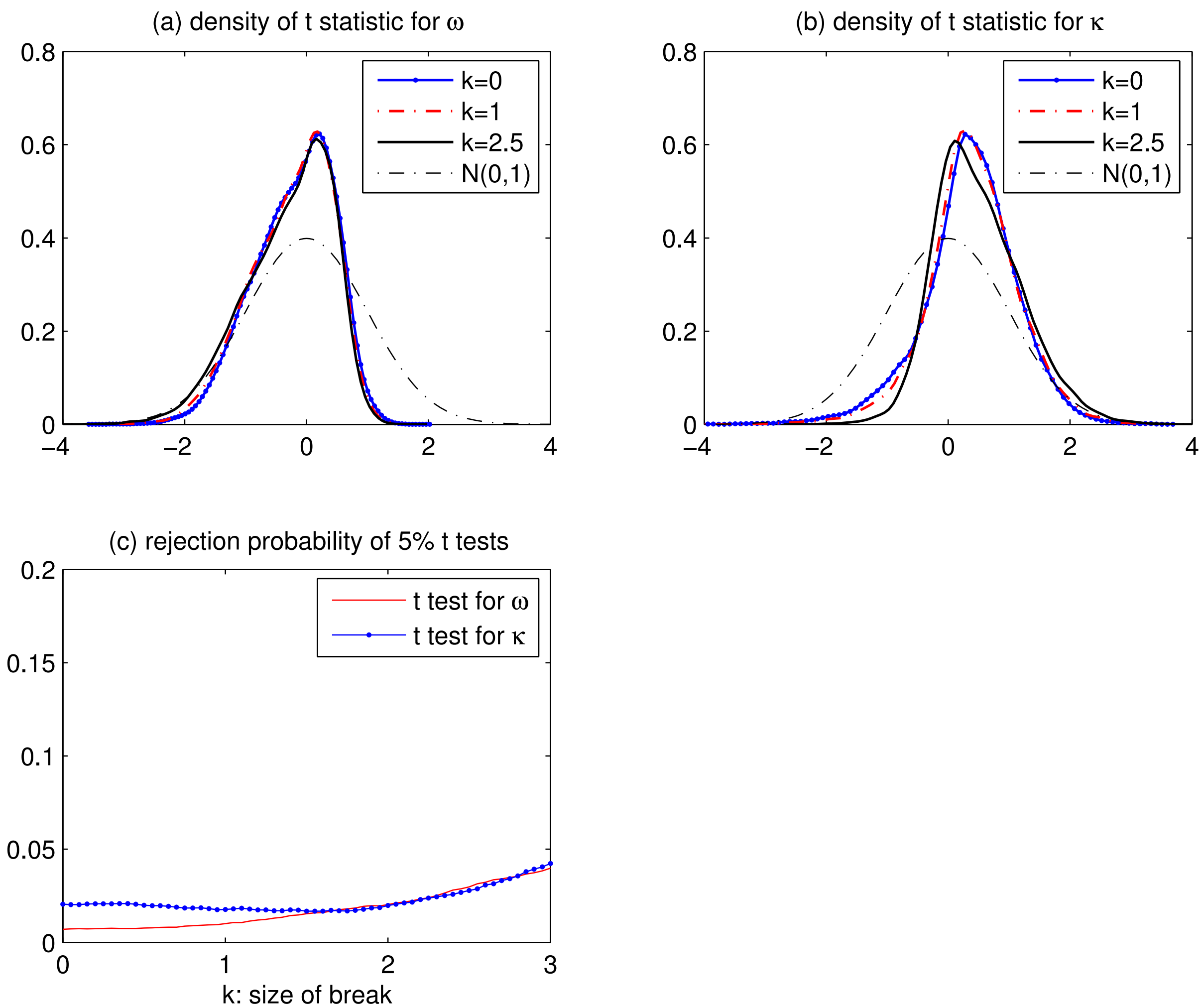
Taylor, J. 1993. Discretion versus policy rules in practice, Carnegie-Rochester conference series on public policy 39, 195 − 214.

Figure 1: Finite sample performance of t test: weak instrument



Notes: Figure 1 displays the small sample behavior of the t test in a scalar regression with *weak* instrument. Panel a plots the small sample (T=100) distribution of the t test when $\lambda = 0$ (no instability), $\lambda = 8$ (moderate instability) and $\lambda = 16$ (very large instability). Panel b plots the empirical size of the t test versus the magnitude of parameter instability. The number of replications is 10,000.

Figure 2: Finite sample performance of t test: weakly identified Euler equation



Notes: Figure 2 displays the small sample (T=100) behavior of the t tests in a forward-looking model (26) featuring rational expectations. Panels a and b plot the small sample distributions of the t tests when the break in the reduced-form process (27) is zero ($K = 0$), small ($K = 1$) and moderate ($K = 2.5$). Panel c plots the empirical sizes of the t tests versus the magnitude of the break in the reduced-from parameters. The number of replications is 20,000.