Exponential Tilting with Weak Instruments: Estimation and Testing

Mehmet Caner
University of Pittsburgh *

September 2003

Abstract

This article analyzes exponential tilting estimator with weak instruments in a nonlinear framework. The limits of these estimators under standard identification assumptions are derived by Imbens, Spady and Johnson (1998) and Kitamura and Stutzer (1997). We obtain the new limits when the instruments are weakly correlated with the moment restrictions. First, we show that Lagrange Multipliers are affected by weak instruments and this results in the inconsistent estimates for the weakly identified parameters. In this paper, we obtain the limits of both Lagrange Multiplier estimates and the estimates of the parameters in moment restrictions. The limit of the estimators of Lagrange Multipliers are no longer normally distributed and depends on the limits of the parameter estimates. In this respect, weak instrument asymptotics are different from standard asymptotics, where the two limits are uncorrelated. This dependence affects the limit of J statistic which is not nuisance parameter free. We suggest a new J statistic which is robust to identification and the dependency problem. The results related to limit of Lagrange Multipliers and J test are new in this literature. The limits of the parameter estimators are also derived and they are asymptotically equivalent to the continuously updating version of GMM in the case of weak instrument asymptotics in Stock and Wright (2000). Tests that are robust to identification problem are also obtained. These are Anderson-Rubin and Kleibergen type of test statistics. The limits are nuisance parameter free and χ^2 distributed. We can also build confidence intervals by inverting these test statistics.

JEL Classification: C2,C4,C5.

Keywords: Lagrange multipliers, weak instruments, information theory.

*Department of Economics, WW Posvar Hall 4S01, Pittsburgh, PA 15260. email: caner@pitt.edu. The author thanks Yuichi Kitamura and Werner Ploberger for comments and suggestions.

1 Introduction

In recent literature, Stock and Wright (2000) have shown that GMM's asymptotic properties change when the instruments are weakly correlated with moment conditions. They show that the limits are not asymptotically normal and the new limits involve nuisance parameters. This weak instrument asymptotics give better results in small samples. Inference that is robust to identification is also pursued by Stock and Wright (2000) and they propose an Anderson-Rubin (1949) like test statistic. The limit is χ^2, with degrees of freedom equal to the number of orthogonality conditions. Kleibergen (2002) also provides an LM-like test statistic which is nuisance parameter free. This statistic has also χ^2 limit with degrees of freedom equal to the number of parameters being tested. This has usually better power properties than the Anderson-Rubin like test when there are many instruments. Confidence intervals are built by inverting these two test statistics. Confidence intervals that are based on LM like statistic of Kleibergen (2002) are never empty whereas Anderson-Rubin based confidence intervals may be empty when the overidentifying restrictions are invalid.

To improve the small sample properties of GMM, Newey and Smith (2001) take a different direction. In a recent article, they propose Generalized Empirical Likelihood Estimators. These include continuous updating, exponential tilting, and empirical likelihood estimators. They compare higher-order asymptotic properties of these estimators and GMM. They find that bias-corrected empirical likelihood is asymptotically efficient relative to the other bias-corrected exponential tilting, continuous updating, and GMM two-step estimators. However, as stated in Imbens, Spady and Johnson (1998) exponential tilting has also desirable properties compared to empirical likelihood. Efficient estimates of implied probabilities are used rather than an inefficient $1/T$ as in empirical likelihood. Influence function of exponential tilting is less affected by perturbation in the Lagrange Multipliers compared to empirical likelihood.

In this paper, we analyze exponential tilting with weak instruments. Imbens, Spady and Johnson (1998) and Kitamura and Stutzer (1997) consider the same model with standard identification conditions. Our paper analyzes the case with weak instruments. We consider the weak instrument setup of Stock and Wright (2000). This is important to applied researchers since we have to see how the asymptotics of exponential tilting may be changing when there is an identification problem. We analyze both estimation and testing issues. We show that Lagrange Multipliers associated with orthogonality conditions are affected by weak instruments problem. This results in the inconsistency of the estimates of the weakly identified parameters. We also derive the limit of estimates Lagrange Multipliers. This is not asymptotically normal and depends on nuisance parameters. This limit also involves the limit of the parameter estimates of the moment restrictions. This is unlike the standard asymptotics covered in the cases of exponential tilting and empirical likelihood of Kitamura and Stutzer (1997), Qin and Lawless (1994), Smith (2000). Since the limit of J statistic in this case is not nuisance free as well, we propose a new J statistic that is robust to identification and dependency problem.

We also derive the limits of the parameter estimates in the moment restrictions. The limits of the estimators are asymptotically equivalent to continuously updating GMM case with weak identification in Stock and Wright (2000).

We propose two tests that are robust to identification problem: Anderson-Rubin and Kleibergen type of test statistics. We show that their limits are χ^2 and nuisance free. Confidence intervals can also be built using these test statistics. We also conduct simulation exercises to analyze the small sample properties of these tests.

We should also mention that our paper is not a simple extension of Stock and Wright (2000) or Kleibergen (2002). We deal with a constrained optimization problem and its theoretical derivations are not obvious from the aforementioned papers. We introduce new proofs for overcoming the obtacles introduced by the constraints and the associated Lagrange Multipliers.

A related paper is by Guggenberger (2003) . He analyzes the Generalized Empirical Likelihood Estimators with weak identification. The proof technique in our paper is entirely different because we show that the problems in orthogonality conditions are reflected in identification problems of concentrated Lagrange Multipliers, since they are shadow prices of these constraints. Compared to Guggenberger (2003) we derive two new results , the first one is the limit of Lagrange Multipliers and the second one is the introduction and derivation of the new limit of J test.

Section 2 introduces the assumptions and the model. Section 3 derives the limits of the estimators. Section 4 considers tests that are robust to identification and confidence intervals. Section 5 conducts simulations. Section 6 concludes. The appendix contains all the proofs. "$\Longrightarrow$" represents weak convergence of random functions on compact parameter space with uniform metric. The existence of the estimator of concentrated Lagrange Multiplier is dealt with in Technical Appendix.

2 The Estimator

Suppose we are given the following moment condition:

$$E_\mu[f(x_t,\theta_0)|Z_t] = 0 \quad t = 1,\cdots,T, \tag{1}$$

where $f(x_t,\theta) = (f_1,\cdots,f_r)'$, r component vector of observable real valued function. $\theta \in \Theta$, Θ is a compact subset of R^d, θ_0 is in the interior of Θ and x_t is the stochastic vector process. E_μ represents the expectation with respect to the probability measure μ. Z_t is the s vector of the instruments. We introduce the notation that helps us understand the estimation. Let γ represent the vector (Lagrange Multiplier) associated with convex optimization problem associated with the constraints in (1) as in Kitamura and Stutzer (1997). Let $\gamma \in R^{rs}$ and $\psi_t(\theta) = f(x_t,\theta)\bigotimes Z_t$,

$rs \geq d$. As in Kitamura and Stutzer (1997)

$$\gamma(\theta) = \arg\min_{\gamma} E_\mu[e^{\gamma' \psi_t(\theta)}], \tag{2}$$

and

$$\theta_0 = \arg\max_{\theta \in \Theta} E_\mu[e^{\gamma(\theta)' \psi_t(\theta)}]. \tag{3}$$

In order to estimate the parameter vector, the exponential tilting estimator in Kitamura and Stutzer (1997) is used. The estimator is

$$(\hat{\theta}_T, \hat{\gamma}_T) = \arg\max_{\theta \in \Theta} \min_{\gamma} \hat{Q}_T(\theta, \gamma), \tag{4}$$

where we set

$$\hat{Q}_T(\theta, \gamma) = \frac{1}{T} \sum_{t=1}^{T} e^{\gamma' \psi_t(\theta)}.$$

From this point let $E[.]$ represent $E_\mu[.]$. We introduce the concept of empirical process $\Psi_T(\theta)$ that is useful for deriving the limit of estimators.

$$\Psi_T(\theta) = T^{-1/2} \sum_{t=1}^{T} \psi_t(\theta) - E\psi_t(\theta)$$

and $\Omega(\theta_1, \theta_2) = \lim_{T \to \infty} E\Psi_T(\theta_1)\Psi_T(\theta_2)'$, and $\tilde{\Psi}_T(\theta) = T^{-1} \sum_{t=1}^{T} \psi_t(\theta)$.

We make the following assumptions:

ASSUMPTIONS:

1.

$$E \sup_{\theta} [\psi_t(\theta) \psi_t(\theta)']$$

is positive definite.

2.

$$E[\sup_{\theta \in \Theta} e^{g' \psi_t(\theta)}] < \infty$$

for all vectors g in a neighborhood of the origin.

3. i) (x_t, Z_t) is iid.

ii)

$$\sup_{\theta \in \Theta} E|\psi_t(\theta)|^{2+\delta} < \infty, \text{for some} \quad \delta > 0.$$

iii)

$$|\psi_t(\theta_1) - \psi_t(\theta_2)| \leq B_t |\theta_1 - \theta_2|$$

where $\lim_{T \to \infty} ET^{-1} \sum_{t=1}^{T} B_t^{2+\delta} < \infty$, for some $\delta > 0$.

4.

$$E\psi_t(\theta) = \frac{m_1(\theta)}{T^{1/2}} + m_2(\beta),$$

where $\theta = (\alpha', \beta')'$, α is $d_1 \times 1$ and β is $d_2 \times 1$ vectors with

i) $m_1(\theta_0) = 0$, $m_1(\theta)$ is continuous in θ and is bounded on Θ.

ii) $m_2(\beta_0) = 0$, $m_2(\beta) \neq 0$, for $\beta \neq \beta_0$, $R(\beta)$ is continuous, $R(\beta_0)$ has full column rank , $R(\beta) = \partial m_2(\beta)/\partial\beta'$ is $rs \times d_2$. Note that by the iid version of the identity in (2.4) of Stock and Wright (2000), we have

$$\frac{m_1(\theta)}{T^{1/2}} = E\psi_t(\alpha, \beta) - E\psi_t(\alpha_0, \beta)$$

and

$$m_2(\beta) = E\psi_t(\alpha_0, \beta).$$

An explanation linking the Stock and Wright (2000) weak identification assumption to our Assumption 4 and justification for Assumption 4 is made in the Remarks after Assumptions.

5. Uniformly in $\theta \in \Theta$,

$$\frac{1}{T}\sum_{t=1}^{T}[\psi_t(\theta) - \tilde{\Psi}_T(\theta)][\psi_t(\theta) - \tilde{\Psi}_T(\theta)]' \xrightarrow{p} \Omega(\theta, \theta).$$

6.

(i).

$$1_{\{\iota'\psi_t(\theta) \geq 0\}} \xrightarrow{a.s.} 0. \qquad for\,all\,t = 1, 2, \cdots, T.$$

(ii).

$$1_{\{\iota'\psi_t(\theta) \leq 0\}} \xrightarrow{a.s.} 0. \qquad for\,all\,t = 1, 2, \cdots, T.$$

Remarks. Assumptions 1-2 are used in the consistency proof and standard in this literature, as shown in Kitamura and Stutzer (1997). Assumption 3 is used in deriving the limits as Assumption B' is used in Stock and Wright (2000). Assumption 4 is the iid version of the weak instrument assumption used in Stock and Wright (2000). In that assumption, α is weakly identified (i.e. in large samples unidentified), and β is identified. This is linked to the constraints in terms of moment equations. Note that Stock and Wright (2000) used the following identity to get the weak identification assumption for the m-dependent random variables:

$$\begin{aligned} ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha,\beta) &= [ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha,\beta) - ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha_0,\beta)] \\ &+ [ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha_0,\beta) - ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha_0,\beta_0)] \\ &+ [ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha_0,\beta_0)]. \end{aligned} \tag{5}$$

Then, Stock and Wright (2000) assumed the following to get the weak identification in α:

$$ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha,\beta) - ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha_0,\beta) = \frac{m_{1T}(\alpha,\beta)}{T^{1/2}}, \tag{6}$$

where $m_{1T}(\theta) \to m_1(\theta)$ uniformly in θ, and they set

$$ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha_0,\beta) - ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha_0,\beta_0) = m_2(\beta). \tag{7}$$

The third term on the right hand side of (5) is zero by the orthogonality conditions. Combining the (6)-(7) in (5) they get as their weak identification Assumption

$$ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha,\beta) = \frac{m_{1T}(\theta)}{T^{1/2}} + m_2(\beta). \tag{8}$$

Our Assumption 4 is the iid version of their Assumption. In our case, the identity is

$$\begin{aligned} E\psi_t(\alpha,\beta) &= E\psi_t(\alpha,\beta) - E\psi_t(\alpha_0,\beta) \\ &+ E\psi_t(\alpha_0,\beta) - E\psi_t(\alpha_0,\beta_0) \\ &+ E\psi_t(\alpha_0,\beta_0). \end{aligned} \tag{9}$$

Then we assume α is unidentified in large samples, through the following similar to (6)

$$E\psi_t(\alpha,\beta) - E\psi_t(\alpha_0,\beta) = \frac{m_1(\alpha,\beta)}{T^{1/2}}, \tag{10}$$

and then define $E\psi_t(\alpha_0,\beta) = m_2(\beta)$ and $E\psi_t(\alpha_0,\beta_0) = 0$. Using these with (10) in (9) we obtain Assumption 4 above.

An existence proof for the solution of (2) and (4) is also provided by using those assumptions.

Assumption 5 is used for consistent estimation of variance covariance matrix. Assummption 6 is auxiliary and used to provide an alternative existence proof for the solution of (2) and (4). We should also note that the estimation results follows when we replace iid assumption with stationary, ergodic and m-dependent data. However this adds a lot of notation with no change in the limits for estimators so we decided to focus on iid case.

3 Asymptotic Theory

We need a result that is helpful in deriving the limits for estimators. The following Lemma shows that the empirical process weakly converges to a Gaussian limit. We have the following result from section 2 of Andrews (1994):

Lemma 1. *Under Assumption 3* ,

$$\Psi_T(\theta) \Longrightarrow \Psi(\theta)$$

where $\Psi(\theta)$ is a Gaussian process, with mean zero and covariance function $\Omega(\theta_1, \theta_2)$.

Lemma 1 is used in the derivation of the limits for the estimators.

Assumption 4 links the moment condition restriction to the sample size; we can link the Lagrange multiplier corresponding to the constraint to the sample size as well. This is relevant in this case since the Lagrange Multipliers are the "shadow prices" of these constraints. So, similar to (6) of Kitamura and Stutzer (1997), we assume

$$\gamma_T(\alpha, \beta) = \arg\min_{\gamma} E[e^{\gamma' \psi_t(\alpha,\beta)}]. \tag{11}$$

So instead of

$$\gamma(\alpha, \beta) = \arg\min_{\gamma} E[e^{\gamma' \psi_t(\alpha,\beta)}]$$

which is used in (2) we use the version in (11). This formulation helps us to link the weak instruments problem in moment conditions to Lagrange Multipliers associated with these.

Set

$$\gamma(\alpha_0, \beta) = \arg\min_{\gamma} E e^{\gamma' \psi_t(\alpha_0,\beta)}.$$

Before the consistency result for the identified parameters β, we need the following Lemma.

Lemma 2. *Under Assumptions 1-4,*

$$\gamma_T(\alpha, \beta) - \gamma(\alpha_0, \beta) \to 0$$

uniformly in $\theta \in \Theta$, where $\theta = (\alpha', \beta')'$.

Note that the concentrated Lagrange Multipliers $\gamma_T(\alpha, \beta)$ corresponds to the orthogonality condition $E\psi_t(\alpha, \beta)$ in $\frac{m_1(\theta)}{T^{1/2}}$ expression in Assumption 4ii, the $\gamma(\alpha_0, \beta)$ corresponds to $E\psi_t(\alpha_0, \beta)$ by Assumption 4ii. In Assumption 4ii, it is assumed that

$$\frac{m_1(\theta)}{T^{1/2}} = E\psi_t(\alpha, \beta) - E\psi_t(\alpha_0, \beta) \to 0.$$

By Lemma 2 we see that the problems in identification in orthogonality conditions are also reflected in Lagrange Multipliers. Lemma 2 is used in the proof of consistency for the identified parameters β.

Now we show that the identified parameter's estimator is consistent. To prove consistency we use the Wald (1949), Wolfowitz (1949) approach used in Kitamura and Stutzer (1997). However we take into account the unidentification of α in large samples and show that only the estimate of the identified parameter is consistent (β). Theorem 1 in this study generalizes Theorem 1 in Kitamura and Stutzer (1997) to the weak instruments case. The major difference in this case is usage of Lemma 2 and Lemma A.1 in the Appendix. Via these lemmata we benefit from the identification problem for Lagrange multipliers.

Theorem 1 . *Under Assumptions 1-4,*

$$\hat{\beta}_T \xrightarrow{p} \beta_0.$$

We need to find the rate of convergence for the identified parameter estimate before the limit laws are established.

Lemma 3. *Under Assumptions 1-5,*

$$T^{1/2}(\hat{\beta}_T - \beta_0) = O_p(1).$$

In the following Theorem instead of $\Omega(\theta,\theta)$ we use $\Omega_{\theta,\theta}$. Let $\hat{\beta}_T(\alpha)$ solve $argmax_{\beta\in\mathcal{B}}\hat{Q}_T(\alpha,\beta,\hat{\gamma}_T(\alpha,\beta))$, and let $\hat{\alpha}$ solve $\arg\max_{\alpha\in A}\hat{Q}_T(\alpha,\hat{\beta}_T(\alpha),\hat{\gamma}_T(\alpha,\hat{\beta}_T(\alpha)))$ and substitute $\hat{\beta}_T = \hat{\beta}(\hat{\alpha})$. We now introduce the notation that is used in Theorem 2. Let $z(\alpha) = \Omega_{\alpha,\beta_0}^{-1/2'}\Psi(\alpha,\beta_0)$, so that $z(\alpha)$ is a mean zero "rs" dimensional Gaussian process with covariance function $Ez(\alpha_1)z(\alpha_2)' = \Omega_{\alpha_1,\beta_0}^{-1/2'}\Omega((\alpha_1',\beta_0')',(\alpha_2',\beta_0')')\Omega_{\alpha_2,\beta_0}^{-1/2}$ and $\mu(\alpha) = \Omega_{\alpha,\beta_0}^{-1/2'}m_1(\alpha,\beta_0)$. Set $F(\alpha) = \Omega_{\alpha,\beta_0}^{-1/2}R(\beta_0)$. For any nonsingular symmetric matrix $C = C^{1/2'}C^{1/2}$ and $C^{-1} = C^{-1/2}C^{-1/2'}$.

Theorem 2 provides limits for exponential tilting estimators in the case of weak instruments benefiting from the empirical process theory. This theorem uses the weak instrument asymptotics for the limit of exponential tilting estimators unlike the standard asymptotics in Kitamura and Stutzer (1997). Using the limit of the objective function in the following Theorem 2i , we establish the limit for estimators in Theorem 2ii.

Theorem 2. *Under Assumptions 1-5,*

i)

$$\begin{aligned}
-2T[\hat{Q}_T(\alpha,\beta_0+b/T^{1/2},\hat{\gamma}_T(\alpha,\beta_0+b/T^{1/2})) \quad &- \quad \hat{Q}_T(\alpha_0,\beta_0,\gamma(\alpha_0,\beta_0))] \\
&\Longrightarrow \quad [\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R(\beta_0)b]'\Omega_{\alpha,\beta_0}^{-1} \\
&\times \quad [\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)+R(\beta_0)b] \\
&\equiv \quad \bar{S}(\alpha,b)
\end{aligned}$$

ii)

$$(\hat{\alpha}_T', T^{1/2}(\hat{\beta}_T-\beta_0)') \xrightarrow{d} (\alpha^{*'}, b^{*'})$$

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

$$S^*(\alpha) = [z(\alpha)+\mu(\alpha)]'[I - F(\alpha)(F(\alpha)'F(\alpha))^{-1}F(\alpha)'][z(\alpha)+\mu(\alpha)]$$

and

$$b^* = -[R(\beta_0)'\Omega_{\alpha^*,\beta_0}^{-1}R(\beta_0)]^{-1}R(\beta_0)'\Omega_{\alpha^*,\beta_0}^{-1/2}[z(\alpha^*)+\mu(\alpha^*)]$$

where $\Omega_{\alpha^*,\beta_0}$ *represents the variance covariance matrix described in Lemma 1 and evaluated at* $\theta = (\alpha^*,\beta_0)$.

Remarks. Theorem 2i provides the limit for the centralized objective function . The limit is the same as in Theorem 1i of Stock and Wright (2000) for the continuously updated GMM case. This can be seen by replacing the limit weight matrix in Theorem 1i of Stock and Wright (2000) by the limit of the efficient weight matrix $\Omega^{-1}_{\alpha,\beta_0}$. We can see why we have the same limit as Stock and Wright (2000). In the proof we first derive an asymptotically equivalent expression for $\hat{\gamma}_T$, by using the first order condition with respect to γ. Then we substitute this into Taylor series expansion of appropriately centered objective function. This centered objective function is shown to be asymptotically equivalent to Continuous Updating GMM objective function.

Theorem 2ii provides the limits for the estimators. It can be seen that these are entirely different from the normal limits by Kitamura and Stutzer (1997) in the case of identified parameters only.

When α is identified ,$\alpha^* = \alpha_0$, then $\mu(\alpha_0) = 0$, since $m_1(\alpha_0, \beta_0) = 0$ by Assumption 3i and $z(\alpha_0) \equiv N(0, I_d)$. In this case we obtain the limits in Theorem 2 or Corollary 1 in Kitamura and Stutzer (1997) for the case of iid data.

$$T^{1/2}(\hat{\beta}_T - \beta_0) \xrightarrow{d} N(0, [R(\beta_0)'\Omega^{-1}_{\alpha_0,\beta_0} R(\beta_0)]^{-1}).$$

The limits in our Theorem 2ii are equivalent to the limits of Continuously Updated GMM estimators in Corollary 4 of Stock and Wright (2000).

When α is completely unidentified (i.e., in small samples as well) $E\psi_t(\alpha, \beta_0) = 0$, for all α, then $m_1(\alpha, \beta_0) = \mu(\alpha) = 0$. So the limits in Theorem 2ii simplify little so that

$$\alpha^* = \arg\min_{\alpha \in A} S^*(\alpha) = z(\alpha)'[I - F(\alpha)(F(\alpha)'F(\alpha))^{-1}F(\alpha)']z(\alpha).$$

However, this cannot be used since α is a nuisance parameter vector and appears in the limit. The case for $\hat{\beta}_T$ does not simplify much.

Theorem 2 can also be extended to strictly stationary, ergodic, m-dependent data (i.e. m is fixed). However, this comes at a substantial cost of notation, and the objective function should be changing in order to allow for time dependency. The results of this section do not change when we use strictly stationary, ergodic and m-dependent data.

In this part, we derive the large sample theory for the estimators of Lagrange Multipliers :$\hat{\gamma}_T = \hat{\gamma}_T(\hat{\theta}_T)$. This gives us an idea whether their distribution is affected by weak instruments. Also, the limit of Lagrange Multipliers affects the J statistic for overidentifying restrictions in exponential tilting so finding that limit is important.

Theorem 3. *Under Assumptions 1-5,*

$$T^{1/2}\hat{\gamma}_T \Longrightarrow \Omega^{-1}_{\alpha^*,\beta_0}[\Psi(\alpha^*, \beta_0) + m_1(\alpha^*, \beta_0) + R(\beta_0)b^*].$$

This shows that estimators of Lagrange Multipliers limit is clearly affected by weak instrument asymptotics. The reason that the limit of Lagrange Multiplier is different than the standard asymptotics stem from the limit behavior of the sample moment $T^{-1/2}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)$. Since Lagrange multiplier is expressed as a functional of that sample moment term via Assumption 4 and Theorem 2 we get a new result compared to the standard asymptotics in exponential tilting estimator. Note that in standard asymptotics in exponential tilting of Kitamura and Stutzer (1997) or empirical likelihood in Smith (2000), $\hat{\theta}_T$ and $\hat{\gamma}_T$ are asymptotically uncorrelated. Here, we clearly see, α^*, and b^* inside the limit for the estimator of Lagrange Multipliers: $\hat{\gamma}_T$. The main reason for this limit in Theorem 3 is the inconsistency of $\hat{\alpha}$. This result for the Lagrange Multipliers in Theorem 3 is new and this affects the limit of J statistic for overidentifying restrictions used in Kitamura and Stutzer (1997). The limit of the J statistic will not be nuisance free in the case of weak instrument asymptotics in Theorems 2 and 3. So we propose a new J statistic which is robust to identification problems in section 4.

Note that when there is identification of all parameters, $\alpha^* = \alpha_0$, the limit in Theorem 3 simplifies. If $\alpha^* = \alpha_0$, $m_1(\alpha_0, \beta_0) = 0$, and $\mu(\alpha_0) = 0$,

$$b^* = -[R(\beta_0)'\Omega^{-1}_{\alpha_0,\beta_0}R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1/2}_{\alpha_0,\beta_0}z(\alpha_0).$$

Then since $z(\alpha_0) = \Omega^{-1/2}_{\alpha_0,\beta_0}\Psi(\alpha_0, \beta_0)$ the limit in Theorem 3 is:

$$\Omega^{-1}_{\alpha_0,\beta_0}\{\Psi(\alpha_0,\beta_0) - R(\beta_0)[R(\beta_0)'\Omega^{-1}_{\alpha_0,\beta_0}R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}_{\alpha_0,\beta_0}\Psi(\alpha_0,\beta_0)\} \equiv N(0, U),$$

where

$$U = \Omega^{-1}_{\alpha_0,\beta_0} - \Omega^{-1}_{\alpha_0,\beta_0}R(\beta_0)[R(\beta_0)'\Omega^{-1}_{\alpha_0,\beta_0}R(\beta_0)]^{-1}R(\beta_0)'\Omega^{-1}_{\alpha_0,\beta_0}.$$

This last expression is the standard limit that is found in empirical likelihood in Qin and Lawless (1994), Smith (2000) and for the exponential tilting in Kitamura and Stutzer (1997).

4 Testing

The limits of estimators depends on nuisance parameters and these estimators are not consistent . The large sample distributions of LR, Wald and LM tests depend on these estimators' limits. So these test statistics limits are not nuisance parameter free. We need test statistics which are asymptotically pivotal.

In this section we introduce two tests for testing the null of $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$. The limits of these are nuisance parameter free even when there is low correlation between instruments and first order conditions introduced as constraints in exponential tilting. The first one is an Anderson-Rubin like test and the second one is an LM-like test. In the case of weak instruments in GMM; Anderson-Rubin like test is introduced by Stock and Wright (2000). This is called S-based test in Stock and Wright (2000). Since we use variance covariance matrix as $S_T(.)$ in this paper, in order not to cause confusion in notation we call this test Anderson-Rubin like test. Here we introduce a similar test in exponential tilting estimator with weak instruments. First we need the following Assumptions :

Assumption T.1. *The following result holds*

$$T^{-1/2}\sum_{t=1}^{T}\psi_t(\theta_0)\xrightarrow{d}N(0,\Omega_{\theta_0,\theta_0})$$

Assumption T.2.

(i)

$$\frac{1}{T}\sum_{t=1}^{T}[\psi_t(\theta_0)-\tilde{\Psi}_T(\theta_0)][\psi_t(\theta_0)-\tilde{\Psi}_T(\theta_0)]'\xrightarrow{p}\Omega_{\theta_0,\theta_0}$$

where $\tilde{\Psi}_T(\theta_0)=\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)$.

(ii).

$$Ee^{g'\psi_t(\theta)}<\infty.$$

where g is in the neighborhood of zero.

These assumptions are used by Stock and Wright (2000) and Kitamura and Stutzer (1997) as well. This is a simple central limit theorem, and variance covariance matrix estimation , these are of course satisfied under more primitive conditions. Assumptions 3,5 provide the following theorem but Assumptions T.1 and T.2 are weaker, so we use them here:

Theorem 4. *Under Assumptions T.1 and T.2, we have the following result:*

$$-2T[log\hat{Q}_T(\theta_0,\hat{\gamma}_T(\theta_0))]\xrightarrow{d}\chi^2_{rs}$$

Therefore the limit is a χ^2 distribution with degrees of freedom equal to the number of orthogonality conditions (rs). In the continuous updating GMM, Theorem 2 of Stock and Wright (2000) used an Anderson-Rubin like test and derive the same limit. This is robust to identification problem.

This Anderson-Rubin like test can be linked to Likelihood Ratio test in Kitamura and Stutzer (1997). The likelihood ratio test for $H_0:\theta=\theta_0$ is :

$$\begin{aligned}LR_T &= 2T[log\hat{Q}_T(\hat{\theta}_T,\hat{\gamma}_T(\hat{\theta}_T))-log\hat{Q}_T(\theta_0,\hat{\gamma}_T(\theta_0))]\\ &= 2Tlog\hat{Q}_T(\hat{\theta}_T,\hat{\gamma}_T(\hat{\theta}_T))+AR_T(\theta_0)\end{aligned}$$

where $AR_T(\theta_0)$ is the Anderson-Rubin like test in Theorem 4 :

$$AR_T(\theta_0)=-2T[log\hat{Q}_T(\theta_0,\hat{\gamma}_T(\theta_0))].$$

As can be seen from Theorem 2 , the limit for the LR test statistic is not nuisance parameter free due to the limit of the first term on the right hand side of the LR expression.

One drawback of the Anderson-Rubin like test is it may reject when the moment restrictions are invalid. To see this point more clearly, we use the J statistic , which is used for testing the validity of moment restrictions in Kitamura and Stutzer (1997) in exponential tilting estimator. Rewrite $AR_T(\theta_0)$ in the following way :

$$AR_T(\theta_0) = LR_T + J_T$$

where $J_T = -2T log \hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T(\hat{\theta}_T))$.

Note that using Theorem 2 we see that J test is not asymptotically pivotal, also this last decomposition above shows that violation of moment restrictions can influence AR test spuriously.

Next we try to setup a test statistic that may result in higher power than the Anderson-Rubin like test. This is similar to Kleibergen's (2002) test statistic for weakly identified GMM . We need the notation below before the following assumption. Denote

$$\bar{p}_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta)}{\partial\theta'}|_{\theta=\theta_0}.$$

Assumption T.3. *The $rs \times 1$ dimensional derivative of $\psi_t(\theta_0)$ with respect to $\theta_i, i = 1, 2, \cdots, d$:*

$$p_{i,t}(\theta_0) = \frac{\partial\psi_t(\theta)}{\partial\theta_i}|_{\theta_0 i}$$

is such that

$$p_{i,t}(\theta_0) - E[p_{i,t}(\theta_0)] = A_i\left(q_{i,t}(\theta_0) - E[q_{i,t}(\theta_0)]\right)$$

with $q_{i,t}(\theta_0) : l_i \times 1$ and A_i a deterministic full rank $rs \times l_i$ dimensional matrix $l_i \leq rs$. The joint limiting behavior of the sums of martingale difference series $\psi_t(\theta_0)$ and $q_{i,t}(\theta_0) - E[q_{i,t}(\theta_0)|I_t]$ satisfy the following Central Limit Theorem:

$$T^{1/2}\begin{bmatrix}\tilde{\Psi}_T(\theta_0)\\ \bar{q}_T(\theta_0)\end{bmatrix} \xrightarrow{d} \begin{bmatrix}\Psi(\theta_0)\\ \Psi_q(\theta_0)\end{bmatrix} \equiv N(0, V(\theta_0))$$

where

$$\bar{q}_T(\theta_0) = T^{-1}\sum_{t=1}^{T} q_t(\theta_0) - E[q_t(\theta_0)].$$

$\bar{q}_T(\theta_0)$ is of dimension $\sum_{i=1}^{d} l_i \times 1$.

$$V(\theta_0) = \begin{bmatrix}\Omega_{\theta_0,\theta_0} & \Omega_{\theta_0,q}\\ \Omega_{q,\theta_0} & \Omega_{q,q}\end{bmatrix}$$

where dimensions of the sub matrices are $\Omega_{\theta_0,\theta_0} : rs \times rs$, $\Omega_{q,\theta_0} : (\sum_{i=1}^{d} l_i) \times rs$, $\Omega_{q,q} : (\sum_{i=1}^{d} l_i) \times (\sum_{i=1}^{d} l_i)$ and $\Omega_{q,\theta_0} = \Omega'_{\theta_0,q}$. Explicitly the sub matrices are

$$\Omega_{q,\theta_0} = \lim_{T\to\infty} E[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}[q_t(\theta_0) - E[q_t(\theta_0)]][\psi_j(\theta_0)]']$$

$$\Omega_{q,q} = \lim_{T\to\infty} E[\frac{1}{T}\sum_{t=1}^{T}\sum_{j=1}^{T}[q_t(\theta_0) - E[q_t(\theta_0)]][q_j(\theta_0) - E[q_j(x_j,\theta_0)]]'$$

Assumption T.4.

$$Ee^{g'\psi(x,\theta_0)} < \infty$$

for all vectors g in the neighborhood of the origin.

Assumption T.5.

$$\frac{1}{T}\sum_{t=1}^{T}[\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)][\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)]' \xrightarrow{p} \Omega_{\theta_0,\theta_0}$$

$$\frac{1}{T}\sum_{t=1}^{T}[\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)][q_t(\theta_0) - \tilde{q}_T(\theta_0)]' \xrightarrow{p} \Omega_{\theta_0,q}$$

$$\frac{1}{T}\sum_{t=1}^{T}[q_t(\theta_0) - \tilde{q}_T(\theta_0)][q_t(\theta_0) - \tilde{q}_T(\theta_0)]' \xrightarrow{p} \Omega_{q,q}$$

where $\tilde{q}_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} q_t(\theta_0)$.

The instruments span only the part of the information set which is relevant to the estimation of θ so $E[q_t(\theta)] = E[q_t(\theta|I_t]]$. Assumption T.3 assumes the existence of a simple Central Limit Theorem for martingale difference sequences. This can be satisfied under weaker conditions. Note that Assumptions 3 and 4 are different from the martingale diffference sequence assumption here, so the test here is valid under martingale difference sequence assumption.

By Assumption T.3, we can also comment on the limit of the derivative for $\tilde{\Psi}_T(\theta_0)$. We see that the limit only holds for that part of the derivative with respect to θ_i which lies in the span of A_i. The degeneracy of the limit can happen when the derivative of $f(x_t, \theta_0)$ in the moment condition in (1) is completely spanned by Z_t. By choosing $A_i = 0$, this can be avoided. In that case $\bar{q}_T(\theta_0)$ does not exist. Another possible degenerate case is when the derivative of several elements of $f(x_t, \theta)$ with respect to θ_i are identical . By specifying appropriate $\bar{q}_T(\theta_0)$ we can avoid this . These are why we need a limit for $\bar{q}_T(\theta_0)$ rather than $\bar{p}_T(\theta_0)$. More on these issues are explained in detail in Kleibergen (2002).

Instead of Assumption 2, we use the weaker Assumption T.4, and instead of Assumption 5 we use Assumption T.5. These assumptions are standard in this literature , as seen in Kitamura and Stutzer (1997), Kleibergen (2002), and Stock and Wright (2000).

As in Kleibergen (2002), we benefit from the first order condition in exponential empirical likelihood :

$$\frac{\partial \hat{Q}_T(\theta, \hat{\gamma}_T(\theta))}{\partial \theta'} = 0.$$

We base this test statistic on an asymptotically equivalent form of the first order condition. The exact first order condition is given in the following equation (12), asymptotically equivalent form is shown in the proof of Theorem 5 as equation (62)).

When evaluated at θ_0 the first order condition is, by (25) of Kitamura and Stutzer (1997);

$$\frac{\partial \hat{Q}_T(\theta, \hat{\gamma}_T(\theta))}{\partial \theta'}|_{\theta_0} = \hat{\gamma}_T(\theta_0)' \bar{D}_T(\theta_0), \tag{12}$$

where

$$\bar{D}_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi(x_t, \theta)}{\partial \theta'}|_{\theta_0} e^{\hat{\gamma}_T(\theta_0)'\psi(x_t,\theta_0)}$$

$\bar{D}_T(\theta_0)$ is of dimension $rs \times d$. Note that (12) is a simplified version of the actual first order condition when we take the partial derivative of the objective function with respect to θ. The algebraic simplifications to reach (12) is shown in the proof of Theorem 2 in Kitamura and Stutzer (1997).

The following Theorem extends the GMM K-statistic in Kleibergen (2002) to exponential tilting estimators. Note that Guggenberger (2003), also considers the K-statistic in generalized empirical likelihood models. The limit in the following Theorem 5 is the same as in Kleibergen (2002).

Theorem 5. *Under Assumptions T.3, T.4, T.5 , the K-statistic for testing $H_0 : \theta = \theta_0$ is*

$$K(\theta_0) = T\tilde{\Psi}_T(\theta_0)' S_T(\theta_0)^{-1/2} P_{S_T(\theta_0)^{-1/2}\bar{D}_T(\theta_0)} S_T(\theta_0)^{-1/2} \tilde{\Psi}_T(\theta_0) \xrightarrow{d} \chi^2_d$$

where

$$P_{S_T(\theta_0)^{-1/2}\bar{D}_T(\theta_0)} = S_T(\theta_0)^{-1/2}\bar{D}_T(\theta_0)[\bar{D}_T(\theta_0)'S_T(\theta_0)^{-1}\bar{D}_T(\theta_0)]^{-1}\bar{D}_T(\theta_0)'S_T(\theta_0)^{-1/2}$$

and $\tilde{\Psi}_T(\theta_0) = T^{-1}\sum_{t=1}^{T}\psi_t(\theta_0)$.

We show that in the proof of Theorem 5, K-statistic in our case is asymptotically equivalent to K-statistic in Kleibergen (2002) for Continuously Updated GMM. The main difference between the K-test in Kleibergen (2002) in the case of continuously updating GMM and the K-test developed here for exponential tilting estimator is : the Jacobian terms of the objective functions. Given Theorem 5 here ; the subtests can be developed easily , simply following sections 3.2 of Kleibergen (2002).

LM test in exponential tilting estimator in Kitamura and Stutzer (1997) has the same form as in K test statistic developed for exponential tilting. The difference between LM and K tests in exponential tilting is: LM in Kitamura and Stutzer (1997) uses the Jacobian estimator $\bar{p}_T(\theta_0)$ whereas K test here uses $\bar{D}_T(\theta_0)$ term. The large sample theory of $\bar{p}_T(\theta_0)$ is not independent of the limit of $\tilde{\Psi}_T(\theta_0)$. So the limit of the LM depends on nuisance parameters. Note that in our K test the Jacobian term $\bar{D}_T(\theta_0)$ is asymptotically independent of the average moment vector $\tilde{\Psi}_T(\theta_0)$, so this results in nuisance parameter free limit.

K statistic in Continuously Updated GMM case of Kleibergen (2002) takes the value of zero when the GMM objective function is at its minimum, maximum and its inflection points. Note that since the K-test that we built does not depend on exact first order condition in (12), it does not take the value zero when the moment restrictions are invalid. If we had instead built our test using (12) at $\hat{\theta}_T$ this test could have taken the value of zero at its maximum point. However using an asymptotically equivalent form we avoid that problem in small samples. In terms of small sample power the K test that is built here is better in that sense compared to an alternative K test which uses (12). By inverting Anderson-Rubin like test statistic and K tests we can have confidence intervals for θ.

We propose a new J statistic for testing overidentifying restrictions in exponential tilting which is robust to identification. This overcomes the difficulties associated with the limits of $\hat{\gamma}_T$, $\hat{\theta}_T$ in a standard J statistic. We can not use the standard J test statistic introduced in Theorem 3 of Kitamura and Stutzer (1997) in exponential tilting case. We think of testing the validity of moment restrictions:

$$E[\psi_t(\theta_0)] = 0,$$

for all $t = 1, 2, \cdots T$. We can test this by testing $\gamma = 0$ as well. However this restriction makes θ unidentified. We benefit from an idea in eqaution (17) of Smith (2000). He introduces a score based J test in empirical likelihood with standard asymptotics. We modify this for our case, since a score based J test uses $\hat{\theta}_T$, which results in limits with nuisance parameters in weak instrument asymptotics. To overcome this problem we evaluate the score of exponential tilting estimator at $\gamma = 0, \theta = \theta_0$, and base our J test on this restricted parameters. By our test we can test the validity of orthogonality restrictions under the null of $H_0 : \theta = \theta_0$. Note that K and J tests in exponential tilting case introduced here are asymptotically independent using Kleibergen (p.9, 2002). This is also clear from the form of the test statistics. So we propose to use J test in the following manner. First test $H_0 : \theta = \theta_0$ by K test. Since this is asymptotically independent from J test, if we do not reject the null we can test the the moment restrictions by J test. Specifically, the score of our objective function $\hat{Q}_T(.)$ in (4) , evaluated at $\gamma = 0, \theta = \theta_0$ is:

$$SC(\theta_0) = \frac{1}{T} \sum_{t=1}^{T} \begin{pmatrix} \psi_t(\theta_0) \\ 0 \end{pmatrix}.$$

Then J statistic is :

$$J(\theta_0) = T \left\{ SC(\theta_0)' \begin{bmatrix} S_T(\theta_0) & \bar{D}_T(\theta_0) \\ \bar{D}_T(\theta_0)' & 0 \end{bmatrix}^{-1} SC(\theta_0) \right\}$$

where $\bar{D}_T(\theta_0)$ is defined in (12).

Theorem 6. *Under Assumptions T.3-T.5,*

$$J(\theta_0) \xrightarrow{d} \chi^2_{rs-d}.$$

5 Simulation

In this section we analyze the size and power of the Anderson-Rubin like test in Theorem 4 and Kleibergen test statistic in Theorem 5. Our Monte Carlo setups use the representative agent intertemporally separable consumption CAPM with CRRA preferences.

5.1 Size

We closely follow the setup in Stock and Wright (2000) for analyzing the size of the various test statistics described in the above paragraph. The "r" Euler equations are (1) with

$$f(X_t, \theta) = \beta \left(\frac{C_{t+1}}{C_t} \right)^{-\alpha} R_{t+1} - \iota_G.$$

where δ is the discount factor , C_t is the consumption, R_t is a $G \times 1$ vector of asset returns and ι_G is a G vector of ones. Then

$$\psi_t(\theta) = [\beta \left(\frac{C_{t+1}}{C_t} \right)^{-\alpha} R_{t+1} - \iota_G] \bigotimes Z_t. \tag{13}$$

where Z_t is a set of "s" instruments. Let $\theta = (\alpha, \beta)'$ and both parameters are bounded. As in Stock and Wright (2000), α is deemed to be weakly identified and β as strongly identified. The design of the Monte Carlo is due to Tauchen (1986), Kocherlakota (1990), Hansen, Heaton and Yaron (1996). We generate the artificial data for (13). These designs are consistent with Euler equations. This is also used and explained by section 4.2 of Stock and Wright (2000) and section 7 of Kleibergen (2002). In order to generate the artificial data a 10^2 dimensional Markov chain is calibrated to approximate a Gaussian VAR(1) fitted to consumption and dividend growth.

$$\begin{pmatrix} c_t \\ di_t \end{pmatrix} = \begin{pmatrix} 0.021 \\ 0.004 \end{pmatrix} + \begin{pmatrix} -0.161 & 0.017 \\ 0.414 & 0.117 \end{pmatrix} \begin{pmatrix} c_{t-1} \\ di_{t-1} \end{pmatrix} + \begin{pmatrix} \epsilon_{c,t} \\ \epsilon_{di,t} \end{pmatrix},$$

where c_t is the log-growth rate of US per capita real annual consumption growth and di_t is the log-growth rate of real annual dividends on the S&P 500. The errors are independently normally distributed with mean zero and $var(\epsilon_{c,t}) = 0.014$, $var(\epsilon_{di,t}) = 0.0012$ and $cor(\epsilon_{c,t}, \epsilon_{di,t}) = 0.43$. Then this VAR(1) generates the asset returns and consumption growth series in this simulation. The VAR(1) coefficient matrix above adjusts the degree of weak instruments, this VAR(1) specified here corresponds to weak instrument specification in Stock and Wright (2000) and Kleibergen (2002).

Assumptions T.1-T.5 are satisfied under more primitive conditions as shown in p.1072 of Stock and Wright (2000), and section 7 of Kleibergen (2002).

Four designs are described in Table 1.

In Table 2 we consider the size of Anderson-Rubin like test that is introduced in Theorem 4. We use $T = 50, 100, 200$. Size of the test is generally very good. For the designs 1,2,3 at $T = 100, 200$,

Table 1: Monte Carlo Design

Design	α_0	β_0	Assets	Instruments
1	1.3	0.97	r_t^s	$1, r_{t-1}^s, c_{t-1}$
2	13.7	1.139	r_t^s	$1, r_{t-1}^s, c_{t-1}$
3	1.3	0.97	r_t^s, r_t^f	$1, c_{t-1}$
4	1.3	0.97	r_t^s, r_t^f	$1, r_{t-1}^s, r_{t-1}^f, c_{t-1}$

Note: $c_t = ln(C_t/C_{t-1}), r_t^f, r_t^s$ represent consumption growth, the risk free rate, and the stock returns respectively.

Table 2: Size at 5% level

	Anderson-Rubin Test				Kleibergen Test			
Designs	1	2	3	4	1	2	3	4
$T = 50$	9.52	11.48	6.16	1.24	3.84	4.52	5.06	5.78
$T = 100$	7.24	8.48	4.26	0.64	5.18	5.36	4.76	8.68
$T = 200$	6.61	6.44	4.54	0.32	3.84	5.74	4.54	9.70

Note: The test statistics are compared to 5% critical values of the limits in Theorem 4 and Theorem 5. These represent the rejection rates for the corresponding nulls in Table 1. For Anderson-Rubin test for designs 1 and 2 $rs = 3$,$\chi_3^2 = 7.81$; for design 3 $rs = 4$, $\chi_4^2 = 9.49$; for design 4 $rs = 8$, $\chi_8^2 = 15.51$. For Kleibergen test for all designs $\chi_2^2 = 5.99$, corresponding to $d = 2$. These are at all 5% levels. We conducted 5000 trials.

the size is around 5-8% at nominal level 5%. The test is conservative for design 4, rejecting less than the nominal level. At $T = 50$, size increases for designs 1 and 2 to 9.5% and 11.5% respectively. Under the same setup , we analyzed the Kleibergen type test statistic in Theorem 5. We used χ_2^2 at 5% nominal level as the critical value (i.e., 5.99) for all designs. The size of the test is very good and near the nominal level even in small samples such as $T = 50$ in designs 1-3, which is better than the performance of the Anderson-Rubin like test. However when the number of orthogonality conditions increases to $rs = 8$ as in Design 4, the size deteriorates in small samples.

5.2 Power

We consider the power of the Anderson-Rubin like test and Kleibergen test. The setup for the power exercise is as follows: for Designs 1, 3, 4 we set $\beta = 0.98$ and varied α (weakly identified parameter) at 1.0, 1.5, 2.0, 2.5. For Design 2 , we set $\beta = 1$ and varied $\alpha = 3.7, 8.7, 18.7, 23.7$. We reported the rejection rates at 5% actual level. So the power is size-adjusted. These finite sample critical values can be obtained from the author on request. The results are reported in Table 3.

As can be seen from Table 3 when we move away from the false null, rejection rates get larger, and the power improves. We have very good power in $T = 100$ in some cases: when $\beta = 0.98$, $\alpha = 1.0$ in designs 3 and 4, the power is around 95%. Power also improves with large samples. Both tests show the same behavior and the results are very similar for Designs 1, 2, and 4. Only in the case of the just identified system do we see the Kleibergen-like test slightly dominate Anderson-Rubin like test. But we think high rejection rates, near 100%, should be interpreted with caution. In the linear moment restriction case, Guggenberger (2002) finds these tests to be inconsistent. Even though we did not analyze this issue in our nonlinear case, this may be true in nonlinear case as well since the main problem is weak identification. These high rates may occur because $S_T(\theta_0)^{-1}$ may be very large in some parameter settings.

6 Conclusion

This paper develops limits for exponential tilting estimators in the case of weak identification. These are very different from the asymptotically normal ones. We also derive test statistics that are robust to identification. Simulations show that Kleibergen type of test statistics have very good small sample properties. An interesting topic may be developing structural change tests within this framework.

Table 3: Size Adjusted Power

	Anderson-Rubin Test				Kleibergen Test			
	Design 1, $\beta = 0.98$				**Design 1,** $\beta = 0.98$			
α	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5
$T = 50$	11.34	6.82	5.62	8.34	12.48	6.34	5.52	7.14
$T = 100$	16.38	7.74	5.96	11.52	20.24	7.4	4.74	10.84
$T = 200$	29.30	8.44	6.54	20.3	38.44	10.34	6.18	22.72
	Design 2, $\beta = 1$				**Design 2,** $\beta = 1$			
α	3.7	8.7	18.7	23.7	3.7	8.7	18.7	23.7
$T = 50$	8.02	10.14	44.72	48.22	8.16	12.56	48.74	51.50
$T = 100$	14.58	34.92	88.74	91.70	14.32	35.96	89.60	90.04
$T = 200$	28.92	78.40	99.96	100.00	31.02	76.20	99.88	99.80
	Design 3, $\beta = 0.98$				**Design 3,** $\beta = 0.98$			
α	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5
$T = 50$	69.86	19.14	6.46	45.18	78.40	21.92	8.54	54.30
$T = 100$	97.04	37.78	9.04	79.22	98.44	42.56	13.00	88.06
$T = 200$	99.96	64.20	11.58	98.06	100.00	74.76	25.48	99.66
	Design 4, $\beta = 0.98$				**Design 4,** $\beta = 0.98$			
α	1.0	1.5	2.0	2.5	1.0	1.5	2.0	2.5
$T = 50$	69.78	17.72	5.24	39.54	66.44	16.10	3.90	23.38
$T = 100$	96.68	37.04	6.81	70.20	96.66	32.40	4.26	60.98
$T = 200$	99.98	66.24	8.72	95.86	100.00	68.30	13.12	96.64

Note: The test statistics are compared to finite sample critical values that are obtained by running the size program in Table 2 . These can be obtained from the author on request. We use the designs in Table 1 with a change in the risk aversion coefficient and fixing the time discount at $\beta = 0.98$ *in Designs 1,3,4 and* $\beta = 1$ *in Design 2. We conducted 5000 trials.*

APPENDIX

Proof of Lemma 2. We want to show that first $Ee^{\gamma'\psi_t(\alpha,\beta)}$ is not locally identifiable at $\alpha = \alpha_0$. Existence of min as a proof is given in Technical Appendix. Now we analyze the expectation term that is minimized to obtain $\gamma_T(\alpha,\beta)$ in (11). First, we can use the mean value theorem for $e^{\gamma'\psi_t(\alpha,\beta)}$ around 0 to have

$$Ee^{\gamma'\psi_t(\alpha,\beta)} = 1 + \gamma' E\psi_t(\alpha,\beta)e^{\bar{\gamma}'\psi_t(\bar{\alpha},\bar{\beta})}, \tag{14}$$

where $\bar{\gamma}$ is in the line segment joining 0, and γ and $\bar{\theta}$ is between θ_0, and θ_u, where $\theta \in \Theta = [\theta_l, \theta_u]$, θ_u represents the upper boundary of the compact Θ. Choosing the lower bound makes no difference in the proof. Technical Lemma 1 in the appendix provides the behavior of $\gamma'\psi_t(\theta)$ which is useful in proving Lemma 2.

We see by Assumptions 2 and 3 ii, and Technical Lemma 1 that uniformly in γ, θ, using Cauchy-Scwartz inequality (detailed proof of (9) is in Technical Appendix)

$$E\psi_t(\theta)(e^{\bar{\gamma}'\psi_t(\bar{\alpha},\bar{\beta})} - 1) \to 0. \tag{15}$$

By using (15) in (14), one has

$$Ee^{\gamma'\psi_t(\alpha,\beta)} - (1 + \gamma' E\psi_t(\alpha,\beta)) \to 0, \tag{16}$$

Then note that by Assumption 4ii, uniformly in $\theta \in \Theta$,

$$E\psi_t(\alpha,\beta) - E\psi_t(\alpha_0,\beta) \to 0. \tag{17}$$

So we can use (17) to rewrite (16):

$$Ee^{\gamma'\psi_t(\alpha,\beta)} - (1 + \gamma' E\psi_t(\alpha_0,\beta)) \to 0. \tag{18}$$

Next we consider

$$\gamma(\alpha_0,\beta) = \arg\min_{\gamma} Ee^{\gamma'\psi_t(\alpha_0,\beta)}. \tag{19}$$

In (19), note that by using the analysis in (14)- (16)

$$Ee^{\gamma'\psi_t(\alpha_0,\beta)} - (1 + \gamma' E\psi_t(\alpha_0,\beta)) \to 0, \tag{20}$$

So clearly by (18),(20), uniformly in $\theta \in \Theta$,

$$Ee^{\gamma'\psi_t(\alpha,\beta)} - Ee^{\gamma'\psi_t(\alpha_0,\beta)} \to 0. \tag{21}$$

Then given Assumption 1 and (21) and the definitions of $\gamma_T(\alpha,\beta)$ and $\gamma(\alpha_0,\beta)$ (equations (11) and (19)) and Lemma 3.2.1 of van der Vaart and Wellner (1996) we have the desired result. **Q.E.D**.

We need the following lemma for the consistency proof.

Lemma A.1. *Under Assumptions 1-4, uniformly in $\theta \in \Theta$,*

$$\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} \xrightarrow{p} Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)}$$

Proof of Lemma A.1. First rewrite, the term on the left-hand side of Lemma A.1,

$$\left(\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} - Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)}\right) + \left(Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)}\right). \tag{22}$$

In (22) the first term can be expressed in the following way:

$$\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} - Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} = \gamma_T(\alpha,\beta)'\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\alpha,\beta) - E\psi_t(\alpha,\beta)\right] + o_p(1)$$

by taking a mean value expansion around 0 for $\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)$ and using the analysis in the proof of Lemma 2 (equations (14)-(16)).

Then by Lemma 2, we have $\gamma_T(\alpha,\beta) \to \gamma(\alpha_0,\beta)$ and $\gamma(\alpha_0,\beta)$ is bounded and away from $\pm\infty$ which can be seen in Technical Appendix. Then Lemma 1 provides uniformly in $(\alpha',\beta')' \in A \times B = \Theta$,

$$\frac{1}{T}\sum_{t=1}^{T}\psi_t(\alpha,\beta) - E\psi_t(\alpha,\beta) = o_p(1).$$

Taking into account the results above, we obtain uniformly in $\theta \in \Theta$,

$$\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} - Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} \xrightarrow{p} 0. \tag{23}$$

Next we need to show the following to end the proof of Lemma A.1:

$$Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} \to Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)}. \tag{24}$$

First use the mean value theorem as used in (14)-(16) to have

$$Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} = 1 + \gamma_T(\alpha,\beta)'E\psi_t(\alpha,\beta) + o(1). \tag{25}$$

and in the same manner

$$Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)} = 1 + \gamma(\alpha_0,\beta)'E\psi_t(\alpha_0,\beta) + o(1), \tag{26}$$

Subtract (26) from (25)

$$\begin{aligned} Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} &- Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)} \\ &= \gamma_T(\alpha,\beta)'E\psi_t(\alpha,\beta) - \gamma(\alpha_0,\beta)'E\psi_t(\alpha_0,\beta) + o(1). \end{aligned} \tag{27}$$

Add and subtract $\gamma(\alpha_0,\beta)'E\psi_t(\alpha,\beta)$ to the right hand side of the equation (27) to have

$$[\gamma_T(\alpha,\beta)-\gamma(\alpha_0,\beta)]'E\psi_t(\alpha,\beta)+\gamma(\alpha_0,\beta)'[E\psi_t(\alpha,\beta)-E\psi_t(\alpha_0,\beta)]+o_(1). \quad (28)$$

Via Lemma 2 and Assumption 4, we obtain uniformly in $\theta\in\Theta$,

$$[\gamma_T(\alpha,\beta)-\gamma(\alpha_0,\beta)]'E\psi_t(\alpha,\beta)\rightarrow 0. \quad (29)$$

Then since $\gamma(\alpha_0,\beta)$ is in the interior of the convex set by Technical Appendix and by Assumption 4, $E\psi_t(\alpha,\beta)-E\psi_t(\alpha_0,\beta)=\frac{m_1(\theta)}{T^{1/2}}\rightarrow 0$ we obtain in (28)

$$\gamma(\alpha_0,\beta)'[E\psi_t(\alpha,\beta)-E\psi_t(\alpha_0,\beta)]\rightarrow 0. \quad (30)$$

Then use (29)-(30) in (27)-(28) to have (24). (24) and (23) gives us the desired result. **Q.E.D.**

Proof of Theorem 1. The first part of the proof proceeds exactly as in equations (13)-(14) of Kitamura and Stutzer (1997). So we repeat the analysis here. Assumption 4ii implies that there is a unique saddle point $(\alpha_0,\beta_0,\gamma(\alpha_0,\beta_0))$ of the function $M\equiv Ee^{\gamma'\psi(\alpha,\beta)}$ which is exactly as in Kitamura and Stutzer (1997), since $P(\alpha_0,\beta_0)=\mu$, $\gamma(\alpha_0,\beta_0)=0$ and the value of the saddle function $M(\alpha_0,\beta_0,\gamma(\alpha_0,\beta_0))=1$. Assumption 4ii also implies, at $\alpha=\alpha_0$ and $\beta\neq\beta_0$, we have equation (13) of Kitamura and Stutzer (1997) :

$$M(\alpha_0,\beta,\gamma(\alpha_0,\beta))<M(\alpha_0,\beta_0,\gamma(\alpha_0,\beta_0))=1. \quad (31)$$

Next proceed exactly as in p.869 of Kitamura and Stutzer (1997) replacing β there with (α_0,β) in our case, using Assumptions 1-3 via Dominated Convergence Theorem, we obtain

$$\lim_{\delta\downarrow 0}E[\sup_{\beta'\in\Gamma(\beta,\delta)}e^{\gamma(\alpha_0,\beta')'\psi_t(\alpha_0,\beta')}]\equiv M(\alpha_0,\beta,\gamma(\alpha_0,\beta)), \quad (32)$$

where $\Gamma(\beta,\delta)$ denotes an open sphere with center β and radius δ. Use the compactness of Θ to cover $\Theta-\Gamma(\beta_0,\delta)$ with a suitably large number H of spheres $\Gamma(\beta_j,\delta_j)$ taking each δ_j small enough so that (31)-(32) provide

$$E[\sup_{\beta'\in\Gamma(\beta,\delta)}e^{\gamma(\alpha_0,\beta')'\psi_t(\alpha_0,\beta')}]<M(\alpha_0,\beta_0,\gamma(\alpha_0,\beta_0))=1$$

We can thus find positive numbers h_j, so that

$$E[\sup_{\beta'\in\Gamma(\beta_j,\delta_j)}e^{\gamma(\alpha_0,\beta')'\psi_t(\alpha_0,\beta')}]=1-2h_j,\quad j=1,2,\cdots,H.$$

If we analyze the parameter space $\Theta-\Gamma(\beta_0,\delta)$ using the equation above,

$$E_\mu[\sup_{\beta'\in\Theta-\Gamma(\beta_0,\delta)}e^{\gamma(\alpha_0,\beta')'\psi_t(\alpha_0,\beta')}]=1-2h, \quad (33)$$

where $h = \min_j h_j$. Use Lemma A.1

$$P[\sup_{\beta' \in \Theta - \Gamma(\beta_0,\delta)} \frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta')'\psi_t(\alpha,\beta')} - Ee^{\gamma(\alpha_0,\beta')\psi_t(\alpha_0,\beta)} > h] < \epsilon/2. \tag{34}$$

Consider (33)-(34) to have

$$P[\sup_{\beta' \in \Theta - \Gamma(\beta_0,\delta)} \frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta')'\psi_t(\alpha,\beta')} > 1-h] < \epsilon/2.$$

By (16)(17) of Kitamura and Stutzer (1997) , noting that $\hat{\gamma}_T(.)$ is defined in (4) and $\gamma_T(.)$ is defined in (11),

$$\frac{1}{T}\sum_{t=1}^{T} e^{\hat{\gamma}_T(\alpha,\beta)\psi_t(\alpha,\beta)} \leq \frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta)\psi_t(\alpha,\beta)} + o_p(1).$$

For large T therefore,

$$P[\sup_{\beta' \in \Theta - \Gamma(\beta_0,\delta)} \frac{1}{T}\sum_{t=1}^{T} e^{\hat{\gamma}_T(\alpha,\beta')'\psi_t(\alpha,\beta')} > 1-h] < \epsilon/2. \tag{35}$$

But from Lemma A.1 and equation (34) it is clear that in the large samples α is not identified and only the consistency of β is relevant. Then we analyze the behavior of the objective function at (α_0, β_0). So by (19)-(20) of Kitamura and Stutzer (1997) we have

$$P[\frac{1}{T}\sum_{t=1}^{T} e^{\hat{\gamma}_T(\alpha_0,\beta_0)'\psi_t(\alpha_0,\beta_0)} < 1-h/2] < \epsilon/2. \tag{36}$$

Then Lemma A.1, and (34)-(36) imply consistency of $\hat{\beta}$. The main difference with the consistency proof for all well identified parameters in Kitamura and Stutzer (1997) is Lemma 2, Lemma A.1 and equation (34). These show that weakly identified parameter vector is not consistent. **Q.E.D**.

Before the rate of convergence proof, we need a result for the variance covariance matrix estimation, and to show consistency of $\hat{\gamma}_T(\hat{\alpha}_T, \hat{\beta}_T)$, (i.e., $\hat{\gamma}_T \xrightarrow{p} 0$).

First, for the variance covariance matrix estimation

$$\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)' = \frac{1}{T}\sum_{t=1}^{T}[\psi_t(\theta) - \tilde{\Psi}_T(\theta)][\psi_t(\theta) - \tilde{\Psi}_T(\theta)]' + \tilde{\Psi}_T(\theta)\tilde{\Psi}_T(\theta)', \tag{37}$$

where $\tilde{\Psi}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)$. Then see that

$$\tilde{\Psi}_T(\theta) = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta) - E\psi_t(\theta) + E\psi_t(\theta).$$

In the above equation by Lemma 1 and Assumption 4, we have

$$\tilde{\Psi}_T(\theta) \xrightarrow{p} m_2(\beta), \tag{38}$$

uniformly in $\theta \in \Theta$.

Use (38) and Assumption 5 in (37) to get

$$\frac{1}{T}\sum_{t=1}^{T}\psi(x_t,\theta)\psi(x_t,\theta)' \xrightarrow{p} \Omega(\theta,\theta)+m_2(\beta)m_2(\beta)'. \tag{39}$$

uniformly in $\theta \in \Theta$.

To save from further notation, set $\hat{\gamma}_T = \hat{\gamma}_T(\hat{\alpha}_T, \hat{\beta}_T)$. The proof for consistency for $\hat{\gamma}_T$ is the same as it is in the well identified case of Kitamura and Stutzer (1997). The proof crucially depends on the usage of an asymptotic bound which is robust to identification (p.871 of Kitamura and Stutzer (1997)). So simply replacing the variance covariance matrix estimation in that proof with (39) here provides the proof. We here show the proof:

By Kitamura and Stutzer (1997,p.870-871) and using (39) here we obtain the following :

$$\hat{Q}_T(\hat{\theta}_T,\gamma)-\hat{Q}_T(\hat{\theta}_T,0)=\frac{1}{T}\gamma'\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)+\gamma'[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)\psi_t(\hat{\theta}_T)']\gamma+o_p(1)$$

Use the definition of $\hat{\gamma}_T(\hat{\theta}_T)$ and the above result to get

$$\hat{\gamma}_T(\hat{\theta}_T)=-\frac{1}{T}S_T(\hat{\theta}_T)^{-1}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)+o_p(1), \tag{40}$$

where

$$S_T(\hat{\theta}_T)=\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)\psi_t(\hat{\theta}_T)'.$$

Note that by (39)

$$S_T(\hat{\theta}_T)=O_p(1). \tag{41}$$

Then, in order to prove the estimate of the Lagrange multiplier converges in probability to zero, since we have (40) and (41), we need to show the following:

$$\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)=O_p(T^{1/2}). \tag{42}$$

Let $g_T=\frac{g}{T^{1/2}}$, where g is arbitrary rs dimensional vector, and note that

$$\begin{aligned}-2Tlog\hat{Q}_T(\hat{\theta}_T,g_T) &\leq -2Tlog\hat{Q}_T(\hat{\theta}_T,\hat{\gamma}_T)\\ &\leq -2Tlog\hat{Q}_T(\theta_0,\hat{\gamma}_T(\theta_0)).\end{aligned} \tag{43}$$

But the last expression is χ^2_{rs} distributed by using the proof of Theorem 3 of Kitamura and Stutzer (1997) or p.871 of Kitamura and Stutzer (1997). (An alternative proof is given in the proof of our

Theorem 4). Next having a Taylor series expansion as in p.871 of Kitamura and Stutzer (1997), we have by (39) and the definition of g_T 's to have the asymptotically negligible term:

$$\begin{aligned}-2Tlog\hat{Q}_T(\hat{\theta}_T, g_T) &= -2Tg_T' \sum \frac{\psi_t(\hat{\theta}_T)}{T} \\ &- \frac{T}{2}g_T' \sum \frac{\psi_t(\hat{\theta}_T)\psi_t(\hat{\theta}_T)'}{T} g_T + o_p(1),\end{aligned}$$

which equals by (41)

$$-2g_T' \sum \psi_t(\hat{\theta}_T) + O_p(1) = \frac{-2}{T^{1/2}} g' \sum \psi_t(\hat{\theta}_T) + O_p(1).$$

But we have via (43), the last equation is asymptotically bounded by χ^2_{rs} so we obtain (42) which shows that $\hat{\gamma}_T \xrightarrow{p} 0$ through (40) and (41).

Proof of Lemma 3. The goal of the proof is to write our objective function in such a way that we can solve the rate of convergence from the proof in Stock and Wright (2000). First we get an asymptotically equivalent expression of $\hat{\gamma}_T$ from its first order condition. Then we substitute this into the Taylor series expansion of the objective function. By appropriately centering this objective function we can show that the problem is asymptotically equivalent to continuous updating GMM case in Stock and Wright (2000). In order to proceed, try to derive asymptotic approximations of our objective function. First we derive an asymptotic approximation for $\hat{\gamma}_T$. As in equation (21) of Kitamura and Stutzer (1997), consider the first order condition concerning $\hat{\gamma}_T$:

$$\sum_{t=1}^{T} \psi_t(\hat{\alpha}_T, \hat{\beta}_T) e^{\hat{\gamma}_T' \psi_t(\hat{\alpha}_T, \hat{\beta}_T)} = 0. \tag{44}$$

Expand $e^{\hat{\gamma}_T' \psi_t(\hat{\alpha}_T, \hat{\beta}_T)}$ in a Taylor series around 0 to get

$$\begin{aligned}\frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\alpha}_T, \hat{\beta}_T) e^{\hat{\gamma}_T' \psi_t(\hat{\alpha}_T, \hat{\beta}_T)} &= \frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\alpha}_T, \hat{\beta}_T) + \frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\alpha}_T, \hat{\beta}_T)\psi_t(\hat{\alpha}_T, \hat{\beta}_T)'\hat{\gamma}_T \\ &+ \frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\alpha}_T, \hat{\beta}_T) \sum_{j=2}^{\infty} \frac{1}{j!}(\hat{\gamma}_T' \psi_t(\hat{\alpha}_T, \hat{\beta}_T))^j \\ &= \frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\alpha}_T, \hat{\beta}_T) + \frac{1}{T}\sum_{t=1}^{T} \psi_t(\hat{\alpha}_T, \hat{\beta}_T)\psi_t(\hat{\alpha}_T, \hat{\beta}_T)'\hat{\gamma}_T \\ &+ O_p(\|\hat{\gamma}_T\|^2).\end{aligned} \tag{45}$$

In the last equality we use

$$\frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta)'(\psi_t(\theta)\psi_t(\theta)')/2 = O_p(1).$$

by Assumption 2 or 3ii via Uniform Law of Large Numbers.

Set

$$\hat{S}_T(\hat{\theta}_T) = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)\psi_t(\hat{\theta}_T)',$$

where $\hat{\theta}_T = (\hat{\alpha}_T, \hat{\beta}_T)$. By (44) and (45)

$$\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T) + \hat{S}_T(\hat{\theta}_T)\hat{\gamma}_T = O_p(\|\hat{\gamma}_T\|^2).$$

Then, use the above equation

$$\hat{S}_T(\hat{\theta}_T)\hat{\gamma}_T = -\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T) + O_p(\|\hat{\gamma}_T\|^2),$$

and proceed

$$T^{1/2}\hat{\gamma}_T = -\left[\hat{S}_T(\hat{\theta}_T)^{-1}\right]\frac{\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)}{T^{1/2}} + O_p(T^{1/2}\|\hat{\gamma}_T\|^2). \tag{46}$$

In (46) see that $O_p(T^{1/2}\|\hat{\gamma}_T\|^2) = o_p(T^{1/2}\|\hat{\gamma}_T\|)$, since $\hat{\gamma}_T = o_p(1)$ which is shown before the proof of Lemma 3. Next, by (41) and (42)

$$\left[\hat{S}_T(\hat{\theta}_T)^{-1}\right]\frac{\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)}{T^{1/2}} = O_p(1)$$

If $T^{1/2}$ is the right rate of convergence for $\hat{\gamma}_T$ then (46) simplifies and

$$T^{1/2}\hat{\gamma}_T = -\left[\hat{S}_T(\hat{\theta}_T)^{-1}\right]\frac{\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)}{T^{1/2}} + o_p(1),$$

and the right hand of the above equation is $O_p(1)$. When we try another rate such as $T^{1/2+\eta}$ where $\eta > 0$, the right hand side of (46) converges to infinity because of (41) and (42). Also when $\eta < 0$, the right hand side terms in(41) converge in probability to zero by (41) and (42).

One important thing we show the rate of convergence for $\hat{\gamma}_T$, $T^{1/2}\hat{\gamma}_T = O_p(1)$. So we establish the following asymptotic approximation

$$\hat{\gamma}_T = -\hat{S}_T(\hat{\theta}_T)^{-1}\frac{\sum_{t=1}^{T}\psi_t(\hat{\theta}_T)}{T} + o_p(T^{-1/2}). \tag{47}$$

Approximate $\hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T)$ to the second order, where the equality after (45) explains the derivation of the order of the remainder:

$$\hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T) = 1 + \frac{1}{T}\sum_{t=1}^{T}\hat{\gamma}_T'\psi_t(\hat{\theta}_T) + \frac{1}{2T}\sum_{t=1}^{T}(\hat{\gamma}_T'\psi_t(\hat{\theta}_T))^2 + O_p(\|\hat{\gamma}_T\|)^3.$$

Then substitute (47) and use $\hat{\gamma}_T = O_p(T^{-1/2})$ in the above equation to get

$$\begin{aligned}
\hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T) &= 1 - \frac{1}{T^2}[\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)]' \hat{S}_T(\hat{\theta}_T)^{-1} [\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)] \\
&+ \frac{1}{2T^2}[\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)]' \hat{S}_T(\hat{\theta}_T)^{-1} [\frac{\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)\psi_t(\hat{\theta}_T)'}{T}] \hat{S}_T(\hat{\theta}_T)^{-1} [\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)] + o_p(T^{-1}) \\
&= 1 - \frac{1}{T^2}[\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)]' \hat{S}_T(\hat{\theta}_T)^{-1} [\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)] \\
&+ \frac{1}{2T^2}[\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)]' \hat{S}_T(\hat{\theta}_T)^{-1} [\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)] \\
&+ o_p(T^{-1}) \\
&= 1 - \frac{1}{2T}[\frac{\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)}{T^{1/2}}]' \hat{S}_T(\hat{\theta}_T)^{-1} [\frac{\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)}{T^{1/2}}] + o_p(T^{-1}). \qquad (48)
\end{aligned}$$

Similarly

$$\hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0)) = 1 - \frac{1}{2T}\left(\frac{\sum_{t=1}^{T} \psi_t(\theta_0)}{T^{1/2}}\right)' S_T(\theta_0)^{-1} \left(\frac{\sum_{t=1}^{T} \psi_t(\theta_0)}{T^{1/2}}\right) + o_p(T^{-1}), \qquad (49)$$

where $S_T(\theta_0) = \frac{\sum_{t=1}^{T} \psi_t(\theta_0)\psi_t(\theta_0)'}{T}$.

Then by (48) and (49)

$$\begin{aligned}
-2T[\hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T(\hat{\theta}_T)) - \hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0))] &= [\frac{\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)}{T^{1/2}}]' \hat{S}_T(\hat{\theta}_T)^{-1} [\frac{\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)}{T^{1/2}}] \\
&- \left(\frac{\sum_{t=1}^{T} \psi_t(\theta_0)}{T^{1/2}}\right)' S_T(\theta_0)^{-1} \left(\frac{\sum_{t=1}^{T} \psi_t(\theta_0)}{T^{1/2}}\right) + o_p(1) \\
&\leq 0. \qquad (50)
\end{aligned}$$

Furthermore, using the empirical process definition and Assumption 4,

$$\frac{\sum_{t=1}^{T} \psi_t(\hat{\theta}_T)}{T^{1/2}} = \frac{1}{T^{1/2}} \sum_{t=1}^{T} (\psi_t(\hat{\theta}_T) - E\psi_t(\hat{\theta}_T)) + T^{1/2} E\psi_t(\hat{\theta}_T). \qquad (51)$$

Use (51) to rewrite the right hand side of the equality in (50):

$$[\Psi_T(\hat{\theta}_T) + m_1(\hat{\theta}_T) + T^{1/2} m_2(\hat{\beta}_T)]' \hat{S}_T(\hat{\theta}_T)^{-1} [\Psi_T(\hat{\theta}_T) + m_1(\hat{\theta}_T) + T^{1/2} m_2(\hat{\beta}_T)] - [\Psi_T(\theta_0)' S_T(\theta_0)^{-1} \Psi_T(\theta_0)] + o_p(1) \leq 0. \qquad (52)$$

(52) has the same structure as equation (A.1) in p.1091 of Stock and Wright (2000) in their rate of convergence proof (except from the $o_p(1)$ term). The only difference is the weight matrices. In the rate of convergence proof in Stock and Wright (2000), they have a generic weight matrix with the assumption that the weight matrix $W_T(\theta) \xrightarrow{p} W(\theta)$ uniformly over θ where both matrices are positive definite. Here, instead of that case, we have a specific $S_T(\theta) = \frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta)\psi_t(\theta)' \xrightarrow{p}$

$\Omega(\theta,\theta) + m_2(\beta)m_2(\beta)'$ by (39) uniformly over $\theta \in \Theta$. This is positive definite by Assumption 2. Using that information and proceeding exactly as in (A.1)-(A.5) of Stock and Wright (2000) provides the result.**Q.E.D**

Proof of Theorem 2i. By (50) we have

$$\begin{aligned} -2T[\hat{Q}_T(\alpha,\beta_0+b/T^{1/2},\hat{\gamma}_T(\alpha,\beta_0+b/T^{1/2})) &- \hat{Q}_T(\alpha_0,\beta_0,\gamma(\alpha_0,\beta_0))] \\ &= \left[\frac{\sum_{t=1}^T \psi_t(\alpha,\beta_0+b/T^{1/2})}{T^{1/2}}\right]' \hat{S}_T(\alpha,\beta_0+b/T^{1/2})^{-1} \\ &\times \left[\frac{\sum_{t=1}^T \psi_t(\alpha,\beta_0+b/T^{1/2})}{T^{1/2}}\right] + o_p(1), \end{aligned} \tag{53}$$

since $\hat{Q}_T(\alpha_0,\beta_0,\gamma(\alpha_0,\beta_0)) = 1$ as $\gamma(\alpha_0,\beta_0) = 0$ in the proof of Theorem 1. Then note that we can obtain a limit for the right-hand side of the above equation as an empirical process in $(\alpha',b')' \in A \times \bar{B}$ where $\bar{B}$ is compact.

So

$$\begin{aligned} T^{-1/2}\sum_{t=1}^T \psi_t(\alpha,\beta_0+b/T^{1/2}) &= \Psi_T(\alpha,\beta_0+b/T^{1/2}) \\ &+ m_1(\alpha,\beta_0+b/T^{1/2}) + T^{1/2}m_2(\beta_0+b/T^{1/2}). \end{aligned}$$

By Lemmata 1, 3, and Assumption 4,

$$\Psi_T(\alpha,\beta_0+b/T^{1/2}) \Longrightarrow \Psi(\alpha,\beta_0). \tag{54}$$

$$m_1(\alpha,\beta_0+b/T^{1/2}) \to m_1(\alpha,\beta_0). \tag{55}$$

$$T^{1/2}m_2(\beta_0+b/T^{1/2}) \to R(\beta_0)b. \tag{56}$$

By Assumption 5 and (39), and benefiting from $m_2(\beta_0) = 0$ in Assumption 4ii,

$$\hat{S}_T(\alpha,\beta_0+b/T^{1/2}) \xrightarrow{p} \Omega_{\alpha,\beta_0}, \tag{57}$$

where Ω_{α,β_0} denote $\Omega(\theta,\theta)$ evaluated at $\theta = (\alpha',\beta_0')'$. All the limits are uniform in $(\alpha',b')' \in A \times \bar{B}$. Use (54)-(57) to have the desired result.**Q.E.D**

Proof of Theorem 2ii. Use Theorem 2i and Lemma 3.2.1 of van der Vaart and Wellner (1996) to have

$$(\hat{\alpha}',T^{1/2}(\hat{\beta}-\beta_0)') \Longrightarrow (\alpha^{*'},b^{*'}) = \arg\min_{(\alpha',b')\in A\times\bar{B}} \bar{S}(\alpha,b)$$

To obtain the concentrated limit $S^*(\alpha)$, fix α, differentiate $\bar{S}(\alpha,b)$ with respect to b, and after some simple algebra

$$\begin{aligned} b^*(\alpha) &= -[R(\beta_0)'\Omega_{\alpha,\beta_0}^{-1}R(\beta_0)]^{-1} \\ &\times R(\beta_0)'\Omega_{\alpha,\beta_0}^{-1}[\Psi(\alpha,\beta_0)+m_1(\alpha,\beta_0)]. \end{aligned}$$

Set $S^*(\alpha) = \bar{S}(\alpha, b^*(\alpha))$ and after some algebra we obtain the expression for $S^*(\alpha)$ in Theorem 2ii as in the proof of Theorem 1ii in Stock and Wright (2000). Then use the continuous mapping theorem and the envelope theorem to have

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

Because $\hat{\beta} = \hat{\beta}(\alpha)$, $T^{1/2}(\hat{\beta} - \beta_0) \Longrightarrow b^*(\alpha^*)$.**Q.E.D.**

Proof of Theorem 3. The consistency of $\hat{\gamma}_T$ is shown in (37)-(43) before the proof of Lemma 3. The rate of convergence is shown in (46)-(47). For the limit, we also benefit from equation (46) and noting that $T^{1/2}\hat{\gamma}_T^2 = o_p(1)$:

$$T^{1/2}\hat{\gamma}_T = -[\hat{S}_T(\hat{\theta}_T)]^{-1} \frac{\sum_{t=1}^T \psi_t(\hat{\theta}_T)}{T^{1/2}} + o_p(1),$$

where

$$\hat{S}_T(\hat{\theta}_T) = \frac{1}{T}\sum_{t=1}^T \psi_t(\hat{\theta}_T)\psi_t(\hat{\theta}_T)'.$$

Then by (39), Theorem 2ii and $m_2(\beta_0) = 0$ (for this last point see Assumption 4):

$$\hat{S}_T(\hat{\theta}_T) \xrightarrow{p} \Omega_{\alpha^*,\beta_0}.$$

Rewrite the following term using the empirical process and Assumption 4:

$$\begin{aligned} T^{-1/2}\sum_{t=1}^T \psi_t(\hat{\theta}_T) &= T^{-1/2}[\sum_{t=1}^T (\psi_t(\hat{\theta}_T) - E\psi_t(\hat{\theta}_T))] + T^{-1/2}\sum_{t=1}^T E\psi_t(\hat{\theta}_T) \\ &= \Psi_T(\hat{\theta}_T) + m_1(\hat{\theta}_T) + T^{1/2} m_2(\hat{\beta}_T). \end{aligned}$$

By Lemma 1, Theorem 2 and Assumption 4, we derive:

$$T^{-1/2}\sum_{t=1}^T \psi_t(\hat{\theta}_T) \Longrightarrow \Psi(\alpha^*, \beta_0) + m_1(\alpha^*, \beta_0) + R(\beta_0)b^*.$$

Use this last result and the limit for $\hat{S}_T(\hat{\theta}_T)$ derived above to have:

$$T^{1/2}\hat{\gamma}_T \Longrightarrow \Omega^{-1}_{\alpha^*,\beta_0}[\Psi(\alpha^*, \beta_0) + m_1(\alpha^*, \beta_0) + R(\beta_0)b^*].$$

Q.E.D.

Proof of Theorem 4. This proof shows that we can derive the limit under weaker conditions than the proof of Theorem 3 in Kitamura and Stutzer (1997). First of all, under Assumption T.2, and using (39) at $\theta = \theta_0$, we have the following:

$$\frac{1}{T}\sum_{t=1}^T \psi_t(\theta_0)\psi_t(\theta_0)' \xrightarrow{p} \Omega(\theta_0, \theta_0)$$

since $m_2(\beta_0) = 0$. Next see that by p.870 of Kitamura and Stutzer (1997) under Assumptions T.1 , T.2,

$$\hat{\gamma}_T(\theta_0) \xrightarrow{p} 0.$$

Then as in the rate of convergence proof in the first order condition for $\hat{\gamma}_T(\theta_0)$ (equation (45)), expand the $e^{\hat{\gamma}_T(\theta_0)'\psi_t(\theta_0)}$ around 0

$$\begin{aligned}
0 &= \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)e^{\hat{\gamma}_T(\theta_0)'\psi_t(\theta_0)} \\
&= \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0) + \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)\psi_t(\theta_0)'\hat{\gamma}_T(\theta_0) \\
&+ \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)\sum_{j=2}^{\infty}\frac{1}{j!}(\hat{\gamma}_T(\theta_0)'\psi_t(\theta_0))^j \\
&= \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0) + \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)\psi_t(\theta_0)'\hat{\gamma}_T(\theta_0) \\
&+ O_p(\|\hat{\gamma}_T(\theta_0)\|^2).
\end{aligned} \tag{58}$$

In the last equality in (58) we use Assumption T.2.

Then by (58) we have

$$S_T(\theta_0)\hat{\gamma}_T(\theta_0) = -\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0) + O_p(\|\hat{\gamma}_T(\theta_0)\|^2)$$

where $S_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)\psi_t(\theta_0)'$. Since $T^{1/2}O_p(\|\hat{\gamma}_T(\theta_0)\|^2) = o_p(1)$, by $\hat{\gamma}_T = O_p(T^{-1/2})$.

$$T^{1/2}\hat{\gamma}_T(\theta_0) = -S_T(\theta_0)^{-1}\frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T^{1/2}} + o_p(1).$$

We therefore derive

$$\hat{\gamma}_T(\theta_0) = -S_T(\theta_0)^{-1}\frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T} + o_p(T^{-1/2}). \tag{59}$$

As in the rate of convergence proof (i.e., (48), (49)) using the approximation of $\hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0))$ to the second order and substituting (59), we have

$$\hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0)) = 1 - \frac{1}{2T}\left(\frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T^{1/2}}\right)' S_T(\theta_0)^{-1}\left(\frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T^{1/2}}\right) + o_p(T^{-1}). \tag{60}$$

Here we show that we can derive (49) under much weaker conditions than in the rate of convergence proof. Using (60), and using the Assumptions T1-T2, under the null

$$\begin{aligned}
-2T[log(\hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0)))] &= \left(\frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T^{1/2}}\right)' S_T(\theta_0)^{-1}\left(\frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T^{1/2}}\right) + o_p(1) \\
&\xrightarrow{d} \chi^2_{rs}.
\end{aligned}$$

Q.E.D

Proof of Theorem 5. Note that by using (59) where we benefited from Assumptions T.1, T.2, we obtain

$$\hat{\gamma}_T(\theta_0) = -S_T(\theta_0)^{-1}\tilde{\Psi}_T(\theta_0) + o_p(T^{-1/2}), \tag{61}$$

where $\tilde{\Psi}_T(\theta_0) = \frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T}$. By comparing Assumptions T.1, with T.3, it is clear that (61) can be obtained using Assumption T.3 as well.

Then our test statistic use the first term on the right hand side of (61) and ignore $o_p(T^{-1/2})$ term. This means that instead of (12) we use the following asymptotically equivalent form to build the K statistic:

$$-S_T(\theta_0)^{-1}\tilde{\Psi}_T(\theta_0)\bar{D}_T(\theta_0). \tag{62}$$

where

$$\bar{D}_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi(x_t,\theta)}{\partial\theta'}|_{\theta_0} e^{\hat{\gamma}_T(\theta_0)'\psi(x_t,\theta_0)}.$$

So the K test is

$$K(\theta_0) = T\tilde{\Psi}_T(\theta_0)'S_T(\theta_0)^{-1/2}P_{S_T(\theta_0)^{-1/2}\bar{D}_T(\theta_0)}S_T(\theta_0)^{-1/2}\tilde{\Psi}_T(\theta_0)$$

where $P_{\{.\}}$ represents the projection matrix with respect to terms in the subscript.

We try to asymptotically approximate $\bar{D}_T(\theta_0)$ term in (12). Consider each $\bar{D}_{Ti}(\theta_0)$, for $i = 1, 2, \cdots, d$.

$$\begin{aligned}\bar{D}_{Ti}(\theta_0) &= \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta)}{\partial\theta_i}|_{\theta_{0i}} \\ &+ \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta)}{\partial\theta_i}|_{\theta_{0i}}\psi_t(\theta_0)'e^{\gamma_t'\psi_t(\theta_0)}\hat{\gamma}_T(\theta_0).\end{aligned} \tag{63}$$

This is obtained by expanding the exponential term in Taylor's series about 0 to first order as in (26) of Kitamura and Stutzer (1997). Taylor's theorem ensures the existence of vectors γ_t. Next substitute (61) in (63)

$$\bar{D}_{Ti}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta)}{\partial\theta_i}|_{\theta_{0i}} - B_T(\theta_0)S_T(\theta_0)^{-1}\tilde{\Psi}_T(\theta_0) + o_p(B_T(\theta_0)T^{-1/2}), \tag{64}$$

where

$$B_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta)}{\partial\theta_i}|_{\theta_{0i}}\psi_t(\theta_0)'e^{\gamma_t'\psi_t(\theta_0)}.$$

We have to show that

$$B_T(\theta_0) \xrightarrow{p} A_i \Omega_{q_i,\theta_0}. \tag{65}$$

where A_i is $rs \times l_i$ and Ω_{q_i,θ_0} is $l_i \times rs$ and is described in Assumption T.3.

By Assumption T.3 and Assumption T.4 and using $\hat{\gamma}_T(\theta_0) \xrightarrow{p} 0$, in combination with Holder's inequality as in (27) of Kitamura and Stutzer (1997), to get

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}\psi_t(\theta_0)' e^{\gamma_t' \psi_t(\theta_0)} - \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}\psi_t(\theta_0)' \xrightarrow{p} 0. \tag{66}$$

Furthermore rewrite

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}\psi_t(\theta_0)' &= \frac{1}{T}\sum_{t=1}^{T}[\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - \frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}}][\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)]' \\
&+ \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}}[\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)]' \\
&+ \frac{1}{T}\sum_{t=1}^{T}[\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - \frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}}]\psi_t(\theta_0)' \\
&+ \frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}}\tilde{\Psi}_T(\theta_0)',
\end{aligned} \tag{67}$$

where $\tilde{\Psi}_T(\theta_0) = T^{-1}\sum_{t=1}^{T}\psi_t(\theta_0)$ and $\frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}} = T^{-1}\sum_{t=1}^{T}\frac{\partial \psi(\theta)}{\partial \theta_i}|_{\theta_{0i}}$.

In (67) on the right hand side, we analyze the second, third, and fourth terms. The second term may be rewritten as:

$$\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}}[\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)]' = \frac{\partial \bar{\psi}(\theta)}{\partial \theta_i}|_{\theta_{0i}}\frac{1}{T}\sum_{t=1}^{T}[\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)]' = 0, \tag{68}$$

where the last equality is obtained by using the definition of $\tilde{\Psi}_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)$. Then the third term is zero in the same manner. For the fourth term

$$\tilde{\Psi}_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0) \xrightarrow{p} 0, \tag{69}$$

by Assumption T.3. Then consider

$$\frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - E(\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}) + \frac{1}{T}\sum_{t=1}^{T}E(\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}). \tag{70}$$

By Assumption T.3

$$\begin{aligned}
&\frac{1}{T}\sum_{t=1}^{T}\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - E(\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}) \\
= &\frac{1}{T}\sum_{t=1}^{T}p_{it}(\theta_0) - E[p_{it}(\theta_0)] \\
\xrightarrow{p} &\ 0
\end{aligned}$$

and by definition or by (15) of Kleibergen (2002)

$$\frac{1}{T}\sum_{t=1}^{T} E(\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}) \rightarrow J(\theta_{i0})$$

Kleibergen (2002) analyzes the situations where $J(\theta_{i0})$ has full rank, weak value as $J(\theta_{i0}) = C_i/T^{1/2}$, C_i being a positive constant vector or $J(\theta_{i0}) = 0$. The results above, in combination with (69)(70) provide for the fourth term in (67)

$$\frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}} \tilde{\Psi}_T(\theta_0)' \xrightarrow{p} 0. \tag{71}$$

Then clearly in (67)

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} \psi_t(\theta_0)' &= \frac{1}{T}\sum_{t=1}^{T}[\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - \frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}}][\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)]' + o_p(1) \\
&= A_i \frac{1}{T}\sum_{t=1}^{T}(q_{i,t}(\theta_0) - \tilde{q}_{i,t}(\theta_0))(\psi_t(\theta_0) - \Psi_T(\theta_0))' \\
&\xrightarrow{p} A_i \Omega(q_i, \theta_0),
\end{aligned} \tag{72}$$

by Assumption T.3 and T.5. $\tilde{q}_{i,t}(\theta_0) = 1/T \sum q_{i,t}(\theta_0)$. Therefore we obtained (65) by the definition of $B_T(\theta_0)$ immediately after (64), and the results (66),(72).

By Assumption T.5

$$S_T(\theta_0) \xrightarrow{p} \Omega_{\theta_0,\theta_0}. \tag{73}$$

Now we simplify $\bar{D}_{T,i}(\theta_0)$ using (72) and (73) in (64):

$$\bar{D}_{T,i}(\theta_0) = p_{T,i}(\theta_0) - A_i \Omega_{q_i,\theta_0} \Omega^{-1}_{\theta_0,\theta_0} \tilde{\Psi}_T(\theta_0) + o_p(T^{-1/2}), \tag{74}$$

where

$$p_{T,i}(\theta_0) = T^{-1}\sum_{t=1}^{T} p_{i,t}(\theta_0) = T^{-1}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}.$$

So,

$$\bar{D}_T(\theta_0) = [\bar{D}_{T,1}(\theta_0), \cdots, \bar{D}_{T,i}(\theta_0), \cdots, \bar{D}_{T,m}(\theta_0)].$$

$\bar{D}_T(\theta_0)$ is asymptotically equivalent to the term in equation (17) of Kleibergen (2002) divided by T. In other words if we denote $\bar{D}_T(\theta_0)$ term in Kleibergen (2002) (i.e., equation (17) in Kleibergen (2002) divided by T) by $\bar{D}_{TK}(\theta_0)$ to differentiate from our corresponding term we have the following relation:

$$\bar{D}_T(\theta_0) = \bar{D}_{TK}(\theta_0) + o_p(T^{-1/2}).$$

Then using this asymptotic equivalence and the order of the asymptotically negligible term, via Assumption T.3, we obtain Lemma 1 and 2 of Kleibergen (2002) by following the exact same steps in the proofs of Lemma 1 and Lemma 2 in Kleibergen (2002). This leads to Theorem 1 in Kleibergen (2002) and hence the desired result.

In terms of notation in K-statistic in Kleibergen (2002) (i.e., equation (22) there) instead of $f_T(\theta_0)/T$ there we have $\bar{\Psi}_T(\theta_0)$ and instead of $V_{ff}(\theta_0)$ there we have asymptotically equivalent $S_T(\theta_0)$. **Q.E.D.**

Proof of Theorem 6. We can rewrite J statistic using $\tilde{\Psi}_T(\theta_0) = T^{-1}\sum_{t=1}^{T}\psi_t(\theta_0)$:

$$\begin{aligned} J(\theta_0) &= T\left\{\tilde{\Psi}_T(\theta_0)'[S_T(\theta_0)^{-1} - S_T(\theta_0)^{-1}\bar{D}_T(\theta_0)(\bar{D}_T(\theta_0)'S_T(\theta_0)^{-1}\bar{D}_T(\theta_0))^{-1}\bar{D}_T(\theta_0)'S_T(\theta_0)^{-1}]\tilde{\Psi}_T(\theta_0)\right\} \\ &= T\left\{\tilde{\Psi}_T(\theta_0)'S_T(\theta_0)^{-1/2}M_{\{S_T(\theta_0)^{-1/2}\bar{D}_T(\theta_0)\}}S_T(\theta_0)^{-1/2}\tilde{\Psi}_T(\theta_0)\right\}, \end{aligned} \tag{75}$$

where $M_{\{S_T(\theta_0)^{-1/2}\bar{D}_T(\theta_0)\}} = I_{rs} - P_{\{S_T(\theta_0)^{-1/2}\bar{D}_T(\theta_0)\}}$. Rank of M is $rs-d$.

Note that

$$\begin{aligned} T^{1/2}S_T(\theta_0)^{-1/2}\tilde{\Psi}_T(\theta_0) &= S_T(\theta_0)^{-1/2}\frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T^{1/2}} \\ &\xrightarrow{d} N(0, I_{rs}). \end{aligned} \tag{76}$$

Using (76) in (75) we have the desired result.**Q.E.D.**

TECHNICAL APPENDIX

Here we can analyze the issue of existence of $\gamma(\theta)$ and $\hat{\gamma}_T(\theta)$. Here we provide proof based on Lemmata A.1-A.3 of Newey and Smith (2001) or Lemmata 7-9 of Guggenberger (2003). Now we can show that alternatively we can replace inf with min.

Technical Lemma 1. *Under Assumptions 1 and 3*

(i)

$$\sup_{\theta\in\Theta,\gamma\in\Gamma_T}|\gamma'\psi_t(\theta)| \xrightarrow{p} 0.$$

(ii)

$$\Gamma_T \subset \hat{\Gamma}_T(\theta)$$

uniformly in $\theta \in \Theta$, *u.w.p.a.1. where* $\Gamma_T = \{\gamma | \|\gamma\| \leq T^{-1/2}c_T^{-1/2}\}$ *and*

$$c_T = T^{-1/2}\sup_{\theta\in\Theta}\|\psi_t(\theta)\|$$

, $\hat{\Gamma}_T(\theta) = \{\gamma \in R^{rs} : \gamma'\psi_t(\theta) \in \mathcal{V}\}$,$\mathcal{V}$ *is an open interval containing zero.*

Proof of Technical Lemma 1. By Assumption 1, since $\sup_\theta P(\psi_t(\theta) \neq 0) > 0$,

$$1_{\{c_T=0\}} \to 0 \quad a.s.$$

Then by Assumptions 1 and 3i,3ii we derive

$$\sup_{\theta\in\Theta}\|\psi_t(\theta)\| = o_p(T^{1/2}).$$

via equation (B3), Lemma 3.2 of Kunsch (p.1227, 1989) and the proof of Lemma 3.2 of Kunsch (1989) . This last proof is also used in empirical likelihood context for time series data in the proof of Theorem 1 in Kitamura (1997). So using the last two results above with c_T definition we have

$$\sup_{\theta\in\Theta,\gamma\in\Gamma_T}|\gamma'\psi_t(\theta)|\leq T^{-1/2}c_T^{-1/2}\sup_{\theta}\|\psi_t(\theta)\|=o_p(1).$$

This last result shows that $\gamma'\psi_t(\theta)\in\Gamma_T$.**Q.E.D.**

Next under Assumptions 1,3,4,5 and Technical Lemma 1 following Lemmata A.2-A.3 of Newey and Smith (2001) or Lemmata 8-9 of Guggenberger (2003) gives

$$\hat{\gamma}_T(\theta)=\arg\min_{\gamma\in\hat{\Gamma}_T(\theta)}\frac{1}{T}\sum_{t=1}^{T}e^{\gamma'\psi_t(\theta)}$$

exists uniformly with probability approaching one. Similar result holds for $\gamma(\theta)$. So we are able to replace "inf" with "min".

An alternative proof for the existence of interior solutions for $\gamma(\theta)$ and $\hat{\gamma}_T(\theta)$ is given. We benefit from the exponential form of the objective function. Define the following hyperplane H for each $t=1,2,\cdots T$.

$$H=\{\psi_t(\theta)\in R^{rs}|\iota'\psi_t(\theta)=0\},$$

where ι is $rs\times 1$ vector of ones. We analyze four possible cases of the values that $\psi_t(\theta)$ may take with respect to the hyperplane H.

Case 1:

$$\iota'\psi_t(\theta)>0,\qquad for all t=1,2,\cdots T.$$

In this case clearly analyzing the objective function in (2) or (4), $\gamma\to-\infty$, since we are in the positive halfspace determined by H, and that value minimizes the objective function at 0.

Case 2:

$$\iota'\psi_t(\theta)<0,\qquad for all t=1,2,\cdots T.$$

In this case our function for all time observations are in the negative halfspace and clearly $\gamma\to\infty$, since with that value the objective function takes the value of zero in (2) and (4) .

Case 3:

$$\iota'\psi_t(\theta)=0,\qquad for all t=1,2,\cdots T.$$

In this case γ can take any value, including arbitrarily large ones.

Note that combination of cases 1-3, do result in extreme values for γ. (i.e. $\iota'\psi_t(\theta)\leq 0$ or $\iota'\psi_t(\theta)\geq 0$ for all t)

So as long as our moment restrictions $\psi_t(\theta)$ is only one side of the hyperplane all the time we can always optimize at very large values of γ approaching ∞.

Case 4: $\psi_t(\theta)$ can take values both in negative and positive parts of halfspaces, As an example $\psi_t(\theta)<0$ for $t=1,2$ and $\psi_t(\theta)>0$, for $t=3,4,\cdots T$. This does not result in arbitrarily large values for γ. The reason is if that were the case the objective function will explode to positive infinity, which will not be optimizing value of the function.

By Assumption 6 we restrict values that $\psi_t(\theta)$ may take along different time periods. In only case 4 the problem is well defined for our purposes.

Proof of equation (9). First by Cauchy-Scwartz inequality and using $\theta = (\alpha', \beta')'$.

$$E\psi_t(\theta)(e^{\bar{\gamma}'\psi_t(\bar{\theta})} - 1) \leq [E(|\psi_t(\theta)|^2)]^{1/2}[E(|e^{\bar{\gamma}'\psi_t(\bar{\theta})} - 1|^2)]^{1/2}$$

By Assumption 3ii, the first square bracket term on the right hand side of the above equation is bounded and finite. For the second term first use the expansion for exponential term

$$e^{\bar{\gamma}'\psi_t(\theta)} - 1 = \bar{\gamma}'\psi_t(\bar{\theta}) + \frac{[\bar{\gamma}'\psi_t(\bar{\theta})]^2}{2} + ...$$

By Technical Lemma 1 we have almost surely

$$e^{\bar{\gamma}'\psi_t(\bar{\theta})} - 1 = \bar{\gamma}'\psi_t(\bar{\theta}) + o(\bar{\gamma}'\psi_t(\bar{\theta})).$$

Then use this last equation and Technical Lemma 1i

$$\begin{aligned} T^{-1}\sum_{t=1}^{T} |e^{\bar{\gamma}'\psi_t(\bar{\theta})} - 1|^2 \quad &\leq \quad (\sup_{\theta,\gamma} |\bar{\gamma}'\psi_t(\bar{\theta})|^2) + o(1) \\ &\xrightarrow{p} \quad 0 \end{aligned}$$

Then via Theorem 18.8i of Davidson (1994)

$$(E|e^{\bar{\gamma}'\psi_t(\bar{\theta})} - 1|^2)^{1/2} \rightarrow 0.$$

This last result combined with the boundednes of the first term in the Cauchy-Schwartz inequality gives the desired result.

REFERENCES

ANDERSON, T.W., AND H.RUBIN (1949): " Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations," *Annals of Mathematical Statistics,* 20, 46-63.

ANDREWS, D.W.K. (1994): "Empirical Process Methods in Econometrics," *Handbook of Econometrics Vol.4* ed. by R.Engle and D.McFadden. Amsterdam:North Holland, 2247-2294.

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

GUGGENBERGER, P. (2003): "Generalized Empirical Likelihood Tests under Partial , Weak and Strong Identification,". Working Paper, Yale University, Department of Economics.

HANSEN, L.P., J. HEATON, AND A. YARON (1996): "Finite Sample Properties of Some Alternative GMM Estimators," *Journal of Business and Economic Statistics,* 14, 262-280.

IMBENS, G.W., R.H. SPADY, AND P. JOHNSON (1998): "Information Theoretic Approaches To Inference In Moment Condition Models" *Econometrica,* 66, 333-357.

KITAMURA, Y. (1997): "Empirical Likelihood Methods with Weakly Dependent Processes" *Annals of Statistics,* 25, 2084-2102.

KITAMURA, Y. AND M. STUTZER (1997): "An Information-Theoretic Alternative to Generalized Method of Moments Estimation" *Econometrica,* 65, 861-875.

KLEIBERGEN, F. (2002): "Testing Parameters in GMM Without Assuming That They Are Identified," Revised version of Tinbergen Institute Discussion Paper , TI 01-067/4.

KOCHERLAKOTA, N. (1990): "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics,* 26, 285-304.

KUNSCH, H.R. (1989): "The jackknife and the bootstrap for general stationary observations," *Annals of Statistics,* 17, 1217-1241.

NEWEY, W.K. AND R.J. SMITH (2001): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,". Working Paper , MIT, Department of Economics.

QIN, J. AND J. LAWLESS (1994): "Empirical Likelihood and General Estimating Equations," *Annals of Statistics*, 22, 300-325.

SMITH, R.J. (2000): "Empirical Likelihood and Inference," in *Application of Differential Geometry to Econometrics*,ed. by M.Salmon and P. Marriot. Cambridge: Cambridge University Press.

STOCK, J.H., AND J.H. WRIGHT (2000): "GMM With Weak Identification," *Econometrica,* 68, 1055-1096.

TAUCHEN, G. (1986): "Statistical Properties of Generalized Method of Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics,* 4, 397-425.

VAN DER VAART, A. W., AND J.A. WELLNER (1996): "Weak Convergence and Empirical Processes," New York and Springer.

WALD, A. (1949): “Note on the Consistency of the Maximum Likelihood Estimate,” *Annals of Mathematical Statistics,* 20, 595-601.

WOLFOWITZ, J. (1949): “On Wald’s Proof of The Consistency of the Maximum Likelihood Estimate,” *Annals of Mathematical Statistics,* 20, 601-603.