Exponential Tilting with Weak Instruments: Estimation and Testing

Mehmet Caner
University of Pittsburgh *

September 2005

Abstract

This article analyzes exponential tilting estimator with weak instruments in a nonlinear framework for the first time in the literature. 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. In this paper, we obtain the limits of both Lagrange Multiplier estimates and the estimates of the parameters in moment restrictions. We show that Lagrange Multipliers are affected by weak instruments and this results in the inconsistent estimates for the weakly identified parameters. 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 the 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 the limit of Lagrange Multipliers and the J test are new in this literature. The limits of the parameter estimators are also derived, and they are asymptotically equivalent to the continuous updating version of GMM in the case of weak instrument asymptotics in Stock and Wright (2000). Tests that are robust to the 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 the recent literature Stock and Wright (2000) have shown that GMM's asymptotic properties change when the instruments are weakly correlated with moment conditions. They also 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 like (1949) 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 (2004) 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 the 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. The influence function of exponential tilting is less affected by perturbation in the Lagrange Multipliers compared to empirical likelihood. Exponential tilting is more robust to misspecification problems.

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. The paper presents the first nonlinear analysis of exponential tilting estimator with weak instruments. We analyze a model with nonlinear moment restrictions. We show that Lagrange Multipliers associated with orthogonality conditions are affected by the weak instruments problem. This results in the inconsistency of the estimates of the weakly identified parameters. We also derive the limit of estimates of Lagrange Multipliers. This is not asymptotically normal and depends on nuisance parameters. This limit depends on the asymptotics of the parameter estimates. 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 parameter free, we propose a new J statistic that is robust to identification and dependency problems.

We also derive the limits of the parameter estimates in the moment restrictions. The limits of the estimators are asymptotically equivalent to weakly identified continuous updating estimator (CUE) limit as in Stock and Wright (2000).

We propose two tests that are robust to the identification problem: Anderson-Rubin and Kleibergen type of test statistics. We show that their limits are χ^2 and nuisance parameter 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.

Our paper is the first paper to analyze nonlinear moment restrictions with weak instruments in an exponential tilting framework. Subsequent to our article Guggenberger and Smith (2003) wrote an illuminating article covering all Generalized Empirical Likelihood Estimators with nonlinear moment restrictions and weak identification. We think that there are some important differences between the two papers.

The differences between Guggenberger and Smith's (2003) paper and this one can be analyzed on two levels. These are issues related to the contents of the papers as well as to the proof methods.

First of all, the limits of the estimators in Guggenberger and Smith (2003) do not include the limit of the estimate of the Lagrange Multiplier ($\hat{\gamma}_T$, Theorem 3 in our paper). We establish in Theorem 3 of our paper that $\hat{\gamma}_T$ is consistent and converges at rate $T^{1/2}$, which is the same rate as in the standard Exponential Tilting Model of Kitamura and Stutzer (1997). This is very surprising since the slopes ($\hat{\beta}$) are not consistent, but the Lagrange Multipliers are consistent, and not affected by the problem of weak identification at the same degree as the slopes. This is an important difference in content between the two papers since this finding can be used as a basis for building overidentifying restrictions tests (J-tests). As a second important difference, Guggenberger and Smith (2003) do not provide any J-test (overidentifying restrictions test) which is crucial in validating implications of economic theory. It is difficult to come up with a J-test since this test uses $\hat{\beta}$, generally which is not a consistent estimate in weak identification literature. In other words, both the form and the limit of the J-test are not easy to find. This is Theorem 6 in the paper and new in this literature.

To see theoretical differences with the Guggenberger and Smith (2003) article, our consistency proof, which includes Lemma 2, Lemma A.1 and Theorem 1, is new in the weak identification literature. Only the last part of the proof of Theorem 1 extends the Wald (1949) and Wolfowitz

(1949) consistency approach to a mixed setup of both weakly and fully identified parameters. The assumption and techniques that Guggenberger and Smith (2003) used in their paper benefit from the approach in Newey and Smith (2004).

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. The existence of the estimator of concentrated Lagrange Multiplier is shown in the Technical Appendix. "$\Longrightarrow$" represents weak convergence of random functions on compact parameter space with uniform metric.

2 The Estimator

Suppose we are given the following moment condition

$$E[\psi(x_t,\theta_0)] = 0 \quad t = 1,\cdots,T, \tag{1}$$

where x_t is the data vector and $\psi(x_t,\theta)$is an $r \times 1$ vector of observable real valued functions. Let $\theta \in \Theta$, Θ is a compact subset of R^d, and θ_0 is in the interior of Θ . $E[]$ represents the expectation taken with respect to the distribution of x_t. We introduce the notation that is helpful for subsequent sections. Let γ represent the $r \times 1$ vector (Lagrange Multiplier) associated with the convex optimization problem associated with the constraints in (1) as in Kitamura and Stutzer (1997). Let $\gamma \in R^r$ and $\psi_t(\theta)$ represent the function $\psi(x_t,\theta)$ from now on, and $r \geq d$.

As in Kitamura and Stutzer (1997)

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

and

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

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

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

We introduce the concept of empirical process $\Psi_T(\theta)$ that is useful for deriving the limit of estimators.

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

with variance-covariance $\Omega_{\theta_1,\theta_2} = E\Psi_T(\theta_1)\Psi_T(\theta_2)'$. Then set $\tilde{\Psi}_T(\theta) = T^{-1}\sum_{t=1}^{T}\psi_t(\theta)$.

We need the following assumptions.

ASSUMPTIONS:

1.$E[\psi_t(\theta)\psi_t(\theta)']$ is positive definite for all $\theta \in \Theta$;

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) is iid;

ii)

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

iii)

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

where $EB_t^{2+\delta} < \infty$, 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 $\beta \neq \beta_0$, $R(\beta) = \partial m_2(\beta)/\partial\beta'$ is $r \times d_2 . R(\beta)$ is continuous, $R(\beta_0)$ has full column rank;

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

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

Remarks. Assumptions 1-2 are used in the consistency proof and are 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. Note that Stock and Wright (2000) used the following as the weak identification assumption for the m-dependent random variables:

$$ET^{-1}\sum_{t=1}^{T}\psi_t(\alpha,\beta) = \frac{m_{1T}(\theta)}{T^{1/2}} + m_2(\beta). \tag{5}$$

Our Assumption 4 is the iid version of their Assumption. So in our case, we use

$$\begin{aligned} E\psi_t(\alpha,\beta) &= E\psi_t(\alpha,\beta) - E\psi_t(\alpha_0,\beta) \\ &+ E\psi_t(\alpha_0,\beta) - E\psi_t(\alpha_0,\beta_0) \\ &+ E\psi_t(\alpha_0,\beta_0). \end{aligned} \tag{6}$$

Then we assume α is unidentified in large samples, and define

$$E\psi_t(\alpha,\beta) - E\psi_t(\alpha_0,\beta) = \frac{m_1(\alpha,\beta)}{T^{1/2}}, \tag{7}$$

and then define $E\psi_t(\alpha_0,\beta) = m_2(\beta)$ and $E\psi_t(\alpha_0,\beta_0) = 0$. Using these with (7) in (6) we obtain Assumption 4 above.

Assumption 5 is used for the consistent estimation of variance covariance matrix.

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 $\Omega_{\theta_1,\theta_2}$.

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

$$\gamma_T(\alpha,\beta) = \arg\min_{\gamma} E[e^{\gamma'\psi_t(\alpha,\beta)}]. \tag{8}$$

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 (8). This formulation helps us to link the weak instruments problem in moment conditions to Lagrange Multipliers associated with these.

Set

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

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 $\theta = (\alpha', \beta')'$.

Remark. Note that the concentrated Lagrange Multipliers $\gamma_T(\alpha, \beta)$ correspond to the orthogonality condition $E\psi_t(\alpha, \beta) = 0$, and $\gamma(\alpha_0, \beta)$ corresponds to $E\psi_t(\alpha_0, \beta) = 0$. 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 for parameters 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 parameters' estimators are consistent. To prove consistency we use the Wald (1949)and 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 parameters are 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 the rate of convergence for the Lagrange Multiplier for the subsequent proofs and results. To save notation set $\hat{\gamma}_T = \hat{\gamma}_T(\hat{\theta}_T)$.

Lemma 3.*Under Assumptions 1-5,*

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

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

Lemma 4.*Under Assumptions 1-5,*

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

Let Ω_{α,β_0} denote $\Omega_{\theta,\theta}$ evaluated at $\theta = (\alpha', \beta_0^{'})'$. 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 $z(\alpha) = \Omega^{-1/2'}_{\alpha,\beta_0}\Psi(\alpha,\beta_0)$, so that $z(\alpha)$ is a mean zero "r" dimensional Gaussian process with covariance function $Ez(\alpha_1)z(\alpha_2)' = \Omega^{-1/2'}_{\alpha_1,\beta_0}\Omega((\alpha_1',\beta_0')',(\alpha_2',\beta_0')')\Omega^{-1/2}_{\alpha_2,\beta_0}$ and $\mu(\alpha) = \Omega^{-1/2'}_{\alpha,\beta_0}m_1(\alpha,\beta_0)$. Set $F(\alpha) = \Omega^{-1/2}_{\alpha,\beta_0}R(\beta_0)$. 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)

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

ii)

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

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

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

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 continuous updating estimator (CUE) 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 explain 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 the centered CUE 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 CUE 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 limit for $\hat{\beta}_T$ does not simplify much.

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,*

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

Note that in standard asymptotics in the 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 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. 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 the S-based test in Stock and Wright (2000). Since we use a variance-covariance matrix as $S_T(.)$ in this paper we call this test Anderson-Rubin like test. (in order not to cause confusion in notation) Here we introduce a similar test with an exponential tilting estimator with weak instruments. First we make the following Assumptions :

Assumption T.1.

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

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 a variance-covariance matrix estimation, which are satisfied under more primitive conditions. Assumptions 3, and 5 prove 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, and under the null of* $H_0 : \theta = \theta_0$*,*

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

Therefore, the limit is a χ^2 distribution with degrees of freedom equal to the number of orthogonality conditions (r). 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 the 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

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

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 the 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 the J test is not asymptotically pivotal, also this last decomposition above shows that violation of the moment restrictions can influence an 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

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

The following assumption is first used in Kleibergen (2002).

Assumption T.3. *The $r\times 1$ 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 $r\times l_i$ dimensional matrix $l_i\le r$. The joint limiting behavior of the sums of martingale difference series $\psi_t(\theta_0)$ and $q_{i,t}(\theta_0)-E[q_{i,t}(\theta_0)]$ 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 $\sum_{i=1}^{d} l_i\times 1$, and

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

where dimensions of the sub matrices are $\Omega_{\theta_0,\theta_0}$: $r\times r$, Σ_{q,θ_0} : $(\sum_{i=1}^{d} l_i)\times r$, $\Sigma_{q,q}$: $(\sum_{i=1}^{d} l_i)\times(\sum_{i=1}^{d} l_i)$ and $\Sigma_{q,\theta_0}=\Sigma'_{\theta_0,q}$. Explicitly the sub matrices are

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

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

Assumption T.4.

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

for all vectors g in the neighborhood of the origin;

Assumption T.5.

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

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

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

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

Even though estimation is using iid DGP in section 3, since testing is done using the true value θ_0, we can weaken the iid assumption to martingale difference sequences as used in Kleibergen (2002). 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 $\psi_t(\theta_0)$ in the moment condition in (1) is completely spanned by the instruments. 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 $\psi_t(\theta)$ with respect to θ_i are identical. By specifying the appropriate $\bar{q}_T(\theta_0)$ we can avoid this. This is why we need a limit for $\bar{q}_T(\theta_0)$ rather than $\bar{p}_T(\theta_0)$. This is explained in greater 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 (9), and asymptotically equivalent form is shown in the proof of Theorem 5 as equation (55).

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

$$\frac{\partial \hat{Q}_T(\theta,\hat{\gamma}_T(\theta))}{\partial\theta'}|_{\theta_0}=\hat{\gamma}_T(\theta_0)'\bar{D}_T(\theta_0), \tag{9}$$

where

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

$\bar{D}_T(\theta_0)$ is of dimension $r \times d$. Note that (9) 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 (9) are 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 and Smith (2003) also consider 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*

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

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

The LM test in the exponential tilting estimator in Kitamura and Stutzer (1997) has the same form as in K-test statistic developed for exponential tilting. The difference between the LM and K tests in exponential tilting is that the LM in Kitamura and Stutzer (1997) uses the Jacobian estimator $\bar{p}_T(\theta_0)$ whereas the K test here uses the $\bar{D}_T(\theta_0)$ 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 a nuisance parameter-free limit.

The K statistic in the continuous updating estimator case of Kleibergen (2002) takes the value of zero when the GMM objective function is at its minimum, maximum, and inflection points. Note that since the K-test that we built does not depend on an exact first order condition in (9), it does not take the value zero when the moment restrictions are invalid. If we had instead built our test using (9) 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 properties power of the K test that is built here is better when compared to an alternative

K test which uses (9). By inverting the Anderson-Rubin like test statistic and the 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 resolves the difficulties associated with the limits of $\hat{\gamma}_T$, $\hat{\theta}_T$ in a standard J statistic. We cannot use the standard J test statistic introduced in Theorem 3 of Kitamura and Stutzer (1997) in the 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 equation (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 resolve this problem we evaluate the score of exponential tilting estimator at $\gamma = 0, \theta = \theta_0$, and base our J test on these restricted parameters. With 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 the 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 the J test in the following manner. First test $H_0 : \theta = \theta_0$ by the K test. Since this is asymptotically independent from the J test, and if we do not reject the null, we can test the the moment restrictions by the 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 (9).

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

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

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

$$\psi_t(\theta) = [\beta \left(\frac{C_{t+1}}{C_t}\right)^{-\alpha} R_{t+1} - \iota_G] \bigotimes Z_t, \tag{10}$$

where δ is the discount factor , C_t is consumption, R_t is a $G \times 1$ vector of asset returns, ι_G is a G vector of ones, and Z_t is a set of 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 is 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 (10). 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. It also corresponds to weak instrument specification in Stock and Wright (2000) and Kleibergen (2002). This simulation setup is used by Stock and Wright (2000) and Kleibergen (2002), who assume martingale difference sequence assumption (Assumption T.3). As they have suggested, no autocorrelation or heteroskedasticity in the moment equations are assumed. Errors are martingale difference sequences at true values (there are no overlapping data). 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 $T = 50, 100, 200$. The size of the test is generally very good. For the designs 1, 2, and 3 at $T = 100, 200$, the size is approximately 5-8% at nominal level 5%. The test is conservative for

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: $c_t = ln(C_t/C_{t-1}), r_t^f, r_t^s$ *represents 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, $r = 3, \chi_3^2 = 7.81$*; for design 3* $r = 4$, $\chi_4^2 = 9.49$*; for design 4* $r = 8$, $\chi_8^2 = 15.51$*. For the Kleibergen test for all designs* $\chi_2^2 = 5.99$*, corresponding to* $d = 2$*. These are at all 5% levels. We conducted 5000 trials.*

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 $r = 8$ as in Design 4, the size deteriorates in small samples.

5.2 Power

We consider the power of the Anderson-Rubin like test and the Kleibergen test. The setup for the power exercise is as follows: for Designs 1, 3, 4 we set $\beta = 0.98$ and vary α (weakly identified parameter) at 1.0, 1.5, 2.0, 2.5. For Design 2, we set $\beta = 1$ and vary $\alpha = 3.7, 8.7, 18.7, 23.7$. We report 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 we see the Kleibergen-like test slightly dominate the Anderson-Rubin like test. But we think high rejection rates, near 100%, should be interpreted with caution. In the linear moment restriction case, Guggenberger and Smith (2003) find these tests to be inconsistent. Even though we did not analyze this issue in our nonlinear case, this may be true in the 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 in 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 same 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, and 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 $\alpha = \alpha_0$. The proof of existence of min is given in the Technical Appendix. Now we analyze the expectation term that is minimized to obtain $\gamma_T(\alpha, \beta)$ in (8). First, we can use the mean value theorem for $e^{\gamma'\psi_t(\alpha,\beta)}$ around $\gamma = 0$ to have

$$Ee^{\gamma'\psi_t(\alpha,\beta)} = 1 + \gamma' E\psi_t(\alpha,\beta) e^{\bar{\gamma}'\psi_t(\alpha,\beta)}, \tag{11}$$

where $\bar{\gamma}$ is in the line segment joining 0 and γ. Technical Lemma 1 in the appendix provides the behavior of $\gamma'\psi_t(\theta)$, which is useful in proving Lemma 2. Note that we use Technical Lemma 1i in the remainder of the proof. Namely, $\sup_{\theta\in\Theta,\gamma\in\Gamma_T} |\gamma'\psi_t(\theta)| \xrightarrow{p} 0$. Γ_T is described in Technical Lemma 1.

We see by Assumptions 2 and 3ii, and Technical Lemma 1 that uniformly in γ, θ, using the Cauchy-Schwartz inequality (detailed proof of (12) is in the Technical Appendix)

$$E\psi_t(\theta)(e^{\bar{\gamma}'\psi_t(\alpha,\beta)} - 1) \to 0. \tag{12}$$

By using (12) in (11), one has

$$Ee^{\gamma'\psi_t(\alpha,\beta)} - (1 + \gamma' E\psi_t(\alpha,\beta)) \to 0. \tag{13}$$

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

$$E\psi_t(\alpha,\beta) - E\psi_t(\alpha_0,\beta) \to 0. \tag{14}$$

So we can use (14) to rewrite (13):

$$Ee^{\gamma'\psi_t(\alpha,\beta)} - (1 + \gamma' E\psi_t(\alpha_0,\beta)) \to 0. \tag{15}$$

Next we consider

$$\gamma(\alpha_0,\beta) = \arg\min_{\gamma} Ee^{\gamma'\psi_t(\alpha_0,\beta)}. \tag{16}$$

In (16), note that by using the analysis in (11)- (13)

$$Ee^{\gamma'\psi_t(\alpha_0,\beta)} - (1 + \gamma' E\psi_t(\alpha_0,\beta)) \to 0. \tag{17}$$

So clearly by (15),(17), uniformly in $\theta \in \Theta$,

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

Then given Assumption 1, (18), the definitions of $\gamma_T(\alpha,\beta)$ and $\gamma(\alpha_0,\beta)$ (equations (8) and (16)), 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,

$$\left(\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} - Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)}\right) + \left(Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)}\right). \tag{19}$$

In (19) 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 (11)-(13)).

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 the Technical Appendix. Via Lemma 1 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 $\theta \in \Theta$,

$$\frac{1}{T}\sum_{t=1}^{T} e^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} - Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} \xrightarrow{p} 0. \tag{20}$$

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

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

First use the mean value theorem as used in (11)-(13) to have

$$Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} = 1 + \gamma_T(\alpha,\beta)'E\psi_t(\alpha,\beta) + o(1). \tag{22}$$

Then in the same manner

$$Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)} = 1 + \gamma(\alpha_0,\beta)'E\psi_t(\alpha_0,\beta) + o(1). \tag{23}$$

Next

$$\begin{aligned} Ee^{\gamma_T(\alpha,\beta)'\psi_t(\alpha,\beta)} &= 1 + \gamma(\alpha_0,\beta)'E\psi_t(\alpha,\beta) + o(1) \\ &= 1 + \gamma(\alpha_0,\beta)'E\psi_t(\alpha_0,\beta) + o(1) \\ &= Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)} + o(1), \end{aligned} \tag{24}$$

where the first equality follows by (22) and Lemma 2, the second equality by (14), and the last equality follows by (23). So (24) proves (21).

(21) and (20) give 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 $M \equiv E e^{\gamma' \psi(\alpha, \beta)}$, which is exactly as in Kitamura and Stutzer (1997), because $P(\alpha_0, \beta_0) = \mu$, $\gamma(\alpha_0, \beta_0) = 0$, 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 $\beta \neq \beta_0$, that we have equation (13) of Kitamura and Stutzer (1997) :

$$M(\alpha_0, \beta, \gamma(\alpha_0, \beta)) < M(\alpha_0, \beta_0, \gamma(\alpha_0, \beta_0)) = 1. \tag{25}$$

Let $\Gamma(\beta, \delta)$ denote an open sphere with center β and radius δ. Next proceed exactly as in p.869 of Kitamura and Stutzer (1997), replacing β there with (α_0, β) as in our case, using Assumptions 1-3 and analyzing the parameter space $\Theta - \Gamma(\beta_0, \delta)$ we obtain

$$E[\sup_{\beta' \in \Theta - \Gamma(\beta_0, \delta)} e^{\gamma(\alpha_0, \beta')' \psi_t(\alpha_0, \beta')}] = 1 - 2h, \tag{26}$$

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

$$P[\sup_{\beta' \in \Theta - \Gamma(\beta_0, \delta)} \frac{1}{T} \sum_{t=1}^{T} e^{\gamma_T(\alpha, \beta')' \psi_t(\alpha, \beta')} - E e^{\gamma(\alpha_0, \beta') \psi_t(\alpha_0, \beta)} > h] < \epsilon/2. \tag{27}$$

Consider (26)-(27) 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 (8),

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

For large T, therefore,

$$P[\sup_{\beta' \in \Theta - \Gamma(\beta_0, \delta)} \frac{1}{T} \sum_{t=1}^{T} e^{\hat{\gamma}_T(\alpha, \beta')' \psi_t(\alpha, \beta')} > 1 - h] < \epsilon/2. \tag{28}$$

But from Lemma A.1 and equation (27), 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

$$P[\frac{1}{T} \sum_{t=1}^{T} e^{\hat{\gamma}_T(\alpha_0, \beta_0)' \psi_t(\alpha_0, \beta_0)} < 1 - h/2] < \epsilon/2. \tag{29}$$

Then Lemma A.1 and (27)-(29) 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 (27). 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

$$\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)' = \frac{1}{T}\sum_{t=1}^{T}[\psi_t(\theta) - \tilde{\Psi}_T(\theta)][\psi_t(\theta) - \tilde{\Psi}_T(\theta)]' + \tilde{\Psi}_T(\theta)\tilde{\Psi}_T(\theta)', \tag{30}$$

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

$$\tilde{\Psi}_T(\theta) \xrightarrow{p} m_2(\beta), \tag{31}$$

uniformly in $\theta \in \Theta$.

Use (31) and Assumption 5 in (30) to get

$$\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)' \xrightarrow{p} \Omega_{\theta,\theta} + m_2(\beta)m_2(\beta)'. \tag{32}$$

uniformly in $\theta \in \Theta$.

To save us 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 (32) here proves the claim. Here we show the proof.

By Kitamura and Stutzer (1997,p.870-871), and using (32) 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

$$\hat{\gamma}_T(\hat{\theta}_T) = -\frac{1}{T}S_T(\hat{\theta}_T)^{-1}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T) + o_p(1), \tag{33}$$

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 (32)

$$S_T(\hat{\theta}_T) = O_p(1). \tag{34}$$

Then, in order to prove that the estimate of the Lagrange multiplier converges in probability to zero, and since we have (33) and (34), we need to show the following

$$\sum_{t=1}^{T} \psi_t(\hat{\theta}_T) = O_p(T^{1/2}). \tag{35}$$

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

$$\begin{aligned} -2Tlog\hat{Q}_T(\hat{\theta}_T, g_T) &\leq -2Tlog\hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T) \\ &\leq -2Tlog\hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0)). \end{aligned} \tag{36}$$

But the last expression is χ^2_r 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 see by (32), 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 (34)

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

Via (36), the last equation is asymptotically bounded by χ^2_r, so we obtain (35) which shows that $\hat{\gamma}_T \xrightarrow{p} 0$ through (33) and (34). Somewhat similar arguments can be found in Newey and Smith (2004).

Proof of Lemma 3. As in equation (21) of Kitamura and Stutzer (1997), consider the first order condition concerning $\hat{\gamma}_T$

$$\sum_{t=1}^{T} \psi_t(\hat{\alpha}_T, \hat{\beta}_T) e^{\hat{\gamma}_T' \psi_t(\hat{\alpha}_T, \hat{\beta}_T)} = 0. \tag{37}$$

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

$$
\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat\alpha_T,\hat\beta_T)e^{\hat\gamma_T'\psi_t(\hat\alpha_T,\hat\beta_T)} &= \frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat\alpha_T,\hat\beta_T)+\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat\alpha_T,\hat\beta_T)\psi_t(\hat\alpha_T,\hat\beta_T)'\hat\gamma_T\\
&+ \frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat\alpha_T,\hat\beta_T)\sum_{j=2}^{\infty}\frac{1}{j!}(\hat\gamma_T'\psi_t(\hat\alpha_T,\hat\beta_T))^j\\
&= \frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat\alpha_T,\hat\beta_T)+\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat\alpha_T,\hat\beta_T)\psi_t(\hat\alpha_T,\hat\beta_T)'\hat\gamma_T\\
&+ O_p(\|\hat\gamma_T\|^2). \qquad (38)
\end{aligned}
$$

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 (37) and (38)

$$\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat\theta_T)+\hat S_T(\hat\theta_T)\hat\gamma_T = O_p(\|\hat\gamma_T\|^2).$$

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

$$T^{1/2}\hat\gamma_T = -\left[\hat S_T(\hat\theta_T)^{-1}\right]\frac{\sum_{t=1}^{T}\psi_t(\hat\theta_T)}{T^{1/2}}+O_p(T^{1/2}\|\hat\gamma_T\|^2). \qquad (39)$$

In (39) 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 (34) and (35)

$$\left[\hat S_T(\hat\theta_T)^{-1}\right]\frac{\sum_{t=1}^{T}\psi_t(\hat\theta_T)}{T^{1/2}} = O_p(1).$$

If $T^{1/2}$ is the right rate of convergence for $\hat\gamma_T$, then (39) simplifies and we have

$$T^{1/2}\hat\gamma_T = -\left[\hat S_T(\hat\theta_T)^{-1}\right]\frac{\sum_{t=1}^{T}\psi_t(\hat\theta_T)}{T^{1/2}}+o_p(1). \qquad (40)$$

It is also clear that 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 (39) converges to infinity because of (34) and (35).

Also when $\eta < 0$, the right hand side terms in (39) converge in probability to zero by (34) and (35). So we establish the rate of convergence as $T^{1/2}$.**Q.E.D**

Proof of Lemma 4. 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 the continuous updating GMM case in Stock and Wright (2000). First, we derive an asymptotic approximation for $\hat{\gamma}_T$. By Lemma 3 and the last equation in the proof of Lemma 3, we establish the following asymptotic approximation

$$\hat{\gamma}_T = -\hat{S}_T(\hat{\theta}_T)^{-1}\frac{\sum_{t=1}^T \psi_t(\hat{\theta}_T)}{T} + o_p(T^{-1/2}). \tag{41}$$

Approximate $\hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T)$ to the second order, where the equality after (38) 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 (41) and use $\hat{\gamma}_T = O_p(T^{-1/2})$ in the above equation to get

$$\begin{aligned}
\hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T) &= 1 - \frac{1}{T^2}[\sum_{t=1}^T \psi_t(\hat{\theta}_T)]'\hat{S}_T(\hat{\theta}_T)^{-1}[\sum_{t=1}^T \psi_t(\hat{\theta}_T)] \\
&+ \frac{1}{2T^2}[\sum_{t=1}^T \psi_t(\hat{\theta}_T)]'\hat{S}_T(\hat{\theta}_T)^{-1}[\frac{\sum_{t=1}^T \psi_t(\hat{\theta}_T)\psi_t(\hat{\theta}_T)'}{T}]\hat{S}_T(\hat{\theta}_T)^{-1}[\sum_{t=1}^T \psi_t(\hat{\theta}_T)] + o_p(T^{-1}) \\
&= 1 - \frac{1}{T^2}[\sum_{t=1}^T \psi_t(\hat{\theta}_T)]'\hat{S}_T(\hat{\theta}_T)^{-1}[\sum_{t=1}^T \psi_t(\hat{\theta}_T)] \\
&+ \frac{1}{2T^2}[\sum_{t=1}^T \psi_t(\hat{\theta}_T)]'\hat{S}_T(\hat{\theta}_T)^{-1}[\sum_{t=1}^T \psi_t(\hat{\theta}_T)] \\
&+ o_p(T^{-1}) \\
&= 1 - \frac{1}{2T}[\frac{\sum_{t=1}^T \psi_t(\hat{\theta}_T)}{T^{1/2}}]'\hat{S}_T(\hat{\theta}_T)^{-1}[\frac{\sum_{t=1}^T \psi_t(\hat{\theta}_T)}{T^{1/2}}] + o_p(T^{-1}). \qquad (42)
\end{aligned}$$

Similarly

$$\hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0)) = 1 - \frac{1}{2T}\left(\frac{\sum_{t=1}^T \psi_t(\theta_0)}{T^{1/2}}\right)' S_T(\theta_0)^{-1}\left(\frac{\sum_{t=1}^T \psi_t(\theta_0)}{T^{1/2}}\right) + o_p(T^{-1}), \tag{43}$$

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

Then by (42) and (43)

$$\begin{aligned} -2T[\hat{Q}_T(\hat{\theta}_T, \hat{\gamma}_T(\hat{\theta}_T)) - \hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0))] &= [\frac{\sum_{t=1}^T \psi_t(\hat{\theta}_T)}{T^{1/2}}]' \hat{S}_T(\hat{\theta}_T)^{-1} [\frac{\sum_{t=1}^T \psi_t(\hat{\theta}_T)}{T^{1/2}}] \\ &- \left(\frac{\sum_{t=1}^T \psi_t(\theta_0)}{T^{1/2}} \right)' S_T(\theta_0)^{-1} \left(\frac{\sum_{t=1}^T \psi_t(\theta_0)}{T^{1/2}} \right) + o_p(1) \\ &\leq 0. \end{aligned} \tag{44}$$

Furthermore, using the empirical process definition and Assumption 4,

$$\frac{\sum_{t=1}^T \psi_t(\hat{\theta}_T)}{T^{1/2}} = \frac{1}{T^{1/2}} \sum_{t=1}^T (\psi_t(\hat{\theta}_T) - E\psi_t(\hat{\theta}_T)) + T^{1/2} E\psi_t(\hat{\theta}_T). \tag{45}$$

Use (45) to rewrite the right hand side of the equality in (44)

$$[\Psi_T(\hat{\theta}_T) + m_1(\hat{\theta}_T) + T^{1/2} m_2(\hat{\beta}_T)]' \hat{S}_T(\hat{\theta}_T)^{-1} [\Psi_T(\hat{\theta}_T) + m_1(\hat{\theta}_T) + T^{1/2} m_2(\hat{\beta}_T)] - [\Psi_T(\theta_0)' S_T(\theta_0)^{-1} \Psi_T(\theta_0)] + o_p(1) \leq 0. \tag{46}$$

Equation (46) 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 $S_T(\theta) = \frac{1}{T} \sum_{t=1}^T \psi_t(\theta) \psi_t(\theta)' \xrightarrow{p} \Omega(\theta, \theta) + m_2(\beta) m_2(\beta)'$ by (32) 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 (44) we have

$$\begin{aligned} -2T[\hat{Q}_T(\alpha, \beta_0 + b/T^{1/2}, \hat{\gamma}_T(\alpha, \beta_0 + b/T^{1/2})) &- \hat{Q}_T(\alpha_0, \beta_0, \gamma(\alpha_0, \beta_0))] \\ &= \left[\frac{\sum_{t=1}^T \psi_t(\alpha, \beta_0 + b/T^{1/2})}{T^{1/2}} \right]' \hat{S}_T(\alpha, \beta_0 + b/T^{1/2})^{-1} \\ &\times \left[\frac{\sum_{t=1}^T \psi_t(\alpha, \beta_0 + b/T^{1/2})}{T^{1/2}} \right] + o_p(1), \end{aligned} \tag{47}$$

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,

$$\Psi_T(\alpha, \beta_0 + b/T^{1/2}) \Longrightarrow \Psi(\alpha, \beta_0), \tag{48}$$

$$m_1(\alpha, \beta_0 + b/T^{1/2}) \to m_1(\alpha, \beta_0), \tag{49}$$

$$T^{1/2} m_2(\beta_0 + b/T^{1/2}) \to R(\beta_0) b. \tag{50}$$

By Assumption 5 and (32), and benefiting from $m_2(\beta_0) = 0$ in Assumption 4ii,

$$\hat{S}_T(\alpha, \beta_0 + b/T^{1/2}) \xrightarrow{p} \Omega_{\alpha,\beta_0}, \tag{51}$$

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 (48)-(51) to have the desired result.**Q.E.D**

Proof of Theorem 2ii. This is immediate from Theorem 1ii of Stock and Wright (2000). **Q.E.D.**

Proof of Theorem 3. The consistency of $\hat{\gamma}_T$ is shown in (30)-(36) before the proof of Lemma 3. The rate of convergence is shown in Lemma 3. Then use (40) to have

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

Then by (32), 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

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

Using this last result and the limit for $\hat{S}_T(\hat{\theta}_T)$ derived above, we have the desired result. **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). Under Assumptions T.1-T.2, following the proof of Lemma 3 very closely, we derive

$$\hat{\gamma}_T(\theta_0) = -S_T(\theta_0)^{-1} \frac{\sum_{t=1}^T \psi_t(\theta_0)}{T} + o_p(T^{-1/2}). \tag{52}$$

Note that the proof of Lemma 3 works under $\theta = \theta_0$, with Assumptions T.1-T.2 instead of Assumptions 1-5.

As in the rate of convergence proof (Lemma 3) using the approximation of $\hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0))$ to the second order and substituting (52), we have

$$\hat{Q}_T(\theta_0, \hat{\gamma}_T(\theta_0)) = 1 - \frac{1}{2T}\left(\frac{\sum_{t=1}^T \psi_t(\theta_0)}{T^{1/2}}\right)' S_T(\theta_0)^{-1}\left(\frac{\sum_{t=1}^T \psi_t(\theta_0)}{T^{1/2}}\right) + o_p(T^{-1}). \tag{53}$$

Here we show that we can derive (43) under much weaker conditions than in the rate of convergence proof. Using (53), and the Assumptions T1-T2, under the null

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

Q.E.D

Proof of Theorem 5. Note that by using (52) where we benefit from Assumptions T.1, T.2, we obtain

$$\hat{\gamma}_T(\theta_0) = -S_T(\theta_0)^{-1}\tilde{\Psi}_T(\theta_0) + o_p(T^{-1/2}), \tag{54}$$

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 (54) can be obtained using Assumption T.3 as well.

Then our test statistic uses the first term on the right-hand side of (54) and ignores $o_p(T^{-1/2})$ term. This means that instead of (9) we use the following asymptotically equivalent form to build the K statistic

$$-S_T(\theta_0)^{-1}\tilde{\Psi}_T(\theta_0)\bar{D}_T(\theta_0), \tag{55}$$

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 the terms in the subscript.

We try to asymptotically approximate $\bar{D}_T(\theta_0)$ term in (9). Consider each $\bar{D}_{Ti}(\theta_0)$ for $i = 1, 2, \cdots, d$.

$$\begin{aligned} \bar{D}_{Ti}(\theta_0) &= \frac{1}{T}\sum_{t=1}^T \frac{\partial\psi_t(\theta)}{\partial\theta_i}|_{\theta_{0i}} \\ &+ \frac{1}{T}\sum_{t=1}^T \frac{\partial\psi_t(\theta)}{\partial\theta_i}|_{\theta_{0i}} \psi_t(\theta_0)' e^{\gamma_t'\psi_t(\theta_0)}\hat{\gamma}_T(\theta_0), \end{aligned} \tag{56}$$

where $\gamma_t \in (0, \hat{\gamma}_T(\theta_0))$. 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 (54) in (56)

$$\bar{D}_{Ti}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - B_T(\theta_0) S_T(\theta_0)^{-1} \tilde{\Psi}_T(\theta_0) + o_p(B_T(\theta_0) T^{-1/2}), \tag{57}$$

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

$$B_T(\theta_0) \xrightarrow{p} A_i \Sigma_{q_i,\theta_0}, \tag{58}$$

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

By Assumptions T.3 and 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), we get

$$\frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} \psi_t(\theta_0)' e^{\gamma_t' \psi_t(\theta_0)} - \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} \psi_t(\theta_0)' \xrightarrow{p} 0. \tag{59}$$

Furthermore, rewrite

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} \psi_t(\theta_0)' &= \frac{1}{T}\sum_{t=1}^{T} [\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - \frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}}][\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)]' \\ &+ \frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}} \tilde{\Psi}_T(\theta_0)', \end{aligned} \tag{60}$$

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 (60) on the right-hand side, we analyze the second term. In the second term consider

$$\tilde{\Psi}_T(\theta_0) = \frac{1}{T}\sum_{t=1}^{T} \psi_t(\theta_0) \xrightarrow{p} 0, \tag{61}$$

by Assumption T.3. Then consider

$$\frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}} = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - E(\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}) + \frac{1}{T}\sum_{t=1}^{T} E(\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}}). \tag{62}$$

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)

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

The results above, in combination with (61) and (62) provide for the second term in (60)

$$\frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}} \tilde{\Psi}_T(\theta_0)' \xrightarrow{p} 0. \tag{63}$$

Then clearly in (60)

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T} \frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} \psi_t(\theta_0)' &= \frac{1}{T}\sum_{t=1}^{T} [\frac{\partial \psi_t(\theta)}{\partial \theta_i}|_{\theta_{0i}} - \frac{\partial \tilde{\Psi}_T(\theta)}{\partial \theta_i}|_{\theta_{0i}}][\psi_t(\theta_0) - \tilde{\Psi}_T(\theta_0)]' + o_p(1) \\ &= A_i \frac{1}{T}\sum_{t=1}^{T} (q_{i,t}(\theta_0) - \tilde{q}_{i,t}(\theta_0))(\psi_t(\theta_0) - \Psi_T(\theta_0))' \\ &\xrightarrow{p} A_i \Sigma_{q_i,\theta_0}, \end{aligned} \tag{64}$$

and by Assumptions T.3 and T.5. $\tilde{q}_{i,t}(\theta_0) = 1/T \sum q_{i,t}(\theta_0)$. Therefore, we obtain (58) by the definition of $B_T(\theta_0)$ immediately after (57), and the results (59), and(64).

By Assumption T.5

$$S_T(\theta_0) \xrightarrow{p} \Omega_{\theta_0,\theta_0}. \tag{65}$$

Now we simplify $\bar{D}_{T,i}(\theta_0)$ using (64) and (65) in (57)

$$\bar{D}_{T,i}(\theta_0) = p_{T,i}(\theta_0) - A_i \Sigma_{q_i,\theta_0} \Omega^{-1}_{\theta_0,\theta_0} \tilde{\Psi}_T(\theta_0) + o_p(T^{-1/2}), \tag{66}$$

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 the $\bar{D}_T(\theta_0)$ 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 the notation in the 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 the J statistic using $\tilde{\Psi}_T(\theta_0) = T^{-1}\sum_{t=1}^{T}\psi_t(\theta_0)$

$$\begin{aligned} J(\theta_0) &= T\left\{\tilde{\Psi}_T(\theta_0)'[S_T(\theta_0)^{-1} - S_T(\theta_0)^{-1}\bar{D}_T(\theta_0)(\bar{D}_T(\theta_0)'S_T(\theta_0)^{-1}\bar{D}_T(\theta_0))^{-1}\bar{D}_T(\theta_0)'S_T(\theta_0)^{-1}]\tilde{\Psi}_T(\theta_0)\right\} \\ &= T\left\{\tilde{\Psi}_T(\theta_0)'S_T(\theta_0)^{-1/2}M_{\{S_T(\theta_0)^{-1/2}\bar{D}_T(\theta_0)\}}S_T(\theta_0)^{-1/2}\tilde{\Psi}_T(\theta_0)\right\}, \end{aligned} \tag{67}$$

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

Note that

$$\begin{aligned} T^{1/2}S_T(\theta_0)^{-1/2}\tilde{\Psi}_T(\theta_0) &= S_T(\theta_0)^{-1/2}\frac{\sum_{t=1}^{T}\psi_t(\theta_0)}{T^{1/2}} \\ &\xrightarrow{d} N(0, I_r). \end{aligned} \tag{68}$$

Using (68) in (67) we have the desired result. Note that by equation (67) asymptotically

$$J(\theta_0) = AR_T(\theta_0) - K(\theta_0),$$

where AR_T and K tests are asymptotically independently distributed with χ^2_r, χ^2_d, respectively. **Q.E.D.**

TECHNICAL APPENDIX

Here we can analyze the issue of existence of $\gamma(\theta)$ and $\hat{\gamma}_T(\theta)$. We provide proof based on Lemmata A.1-A.3 of Newey and Smith (2004) or Lemmata A7-A9 of Guggenberger and Smith (2003). Now we can show that alternatively we can replace inf with min.

Technical Lemma 1. *Under Assumptions 3i,ii*

(i)

$$\max_{1\le t\le T}\sup_{\theta\in\Theta,\gamma\in\Gamma_T}|\gamma'\psi_t(\theta)| \xrightarrow{p} 0,$$

(ii)

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

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

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

$\hat{\Gamma}_T(\theta) = \{\gamma \in R^r : \gamma'\psi_t(\theta) \in \mathcal{V}, t = 1, \cdots, T\}$ *,*$\mathcal{V}$ *is an open interval containing zero.*

Proof of Technical Lemma 1. First, by Lemma D.2 of Kitamura, Tripathi, and Ahn (2004) under Assumptions 3i,ii

$$\max_{1\le t\le T}\sup_{\theta\in\Theta}\|\psi_t(\theta)\| = o_p(T^{1/(2+\delta)}), \tag{69}$$

for some $\delta > 0$. Then, without losing any generality as in Lemma A.7 of Guggenberger and Smith (2003), we assume $c_T \neq 0$. Then

$$c_T = T^{-1/2} \max_t \sup_\theta \|\psi_t(\theta)\| \xrightarrow{p} 0,$$

by (69). So $c_T = o_p(1)$. Then by c_T definition

$$\begin{aligned}
\max_t \sup_{\theta\in\Theta,\gamma\in\Gamma_T} |\gamma'\psi_t(\theta)| &\leq T^{-1/2} c_T^{-1/2} \max_t \sup_\theta \|\psi_t(\theta)\| \\
&= T^{-1/2} c_T^{-1/2} c_T T^{1/2} \\
&= c_T^{1/2} \xrightarrow{p} 0,
\end{aligned}$$

which immediately implies part ii of Technical Lemma 1. **Q.E.D.**

Next under Assumptions 1, 3, 4, 5, and Technical Lemma 1 following Lemmata A.2-A.3 of Newey and Smith (2004) or Lemmata 8-9 of Guggenberger and Smith (2003)

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

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

Proof of equation (12). First, by the Cauchy-Scwartz inequality and using $\theta = (\alpha', \beta')'$

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

By Assumption 3ii, the first square bracketed 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

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

By Technical Lemma 1 we have

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

Then use this last equation and Technical Lemma 1i

$$\begin{aligned}
|e^{\bar{\gamma}'\psi_t(\theta)} - 1|^2 &\leq (\sup_{\theta,\gamma} |\bar{\gamma}'\psi_t(\theta)|^2) + o(1) \\
&\xrightarrow{p} 0.
\end{aligned}$$

Then via Theorem 18.8i of Davidson (1994)

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

This last result combined with the boundedness 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. AND R.J. SMITH (2003): "Generalized Empirical Likelihood Tests under Partial , Weak and Strong Identification,". Working Paper, UCLA , 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.

KITAMURA, Y., G. TRIPATHI, WITH H.T.AHN (2004): "Empirical Likelihood Based Inference in Conditional Moment Restriction Model" *Econometrica*, Forthcoming.

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 (2004): "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators,". Econometrica, 72, 219-255.

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.