Exponential Tilting with Weak Instruments: Estimation and Testing

Mehmet Caner
North Carolina State University *

January 2008

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). There are also papers by Guggenberger and Smith (2005), and Otsu (2006) in a similar context to our paper. They derive the limits of generalized empirical likelihood estimators under weak identification. Our paper differs from them in the context of consistency proof. 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. We also conduct a simulation study where we compare empirical likelihood and continuous updating based tests with exponential tilting based ones. The designs involve GARCH (1,1) and contaminated structural errors. We find that exponential tilting based Kleibergen test has the best size among these competitors.

JEL Classification: C2,C4,C5.
Keywords: Lagrange multipliers, weak instruments, information theory.

*Department of Economics, 4168 Nelson Hall, College of Management, Raleigh, NC 27695. email: mcaner@ncsu.edu.

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. The weak instrument asymptotics provides 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 (2005) 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 (2005) are never empty, whereas Anderson-Rubin based confidence intervals may be empty when the overidentifying restrictions are invalid. Recently Caner (2007) has developed boundedly pivotal structural change tests in weakly identified models with nonlinear moment restrictions.

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 consistency and testing issues. We analyze a model with nonlinear moment restrictions.

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. Our simulation involves weak instruments setup with two different designs. In the first setup, structural errors are GARCH (1,1) and in the second one there is a contaminated error distribution. We compare the tests with empirical likelihood and continuous updating based ones. We find that exponential tilting based Kleibergen test has the best size.

In a related paper, Guggenberger (2003), considers Generalized Empirical Likelihood Estimators (GEL) in weakly identified linear models. A subsequent paper by Guggenberger and Smith (2005) generalize Guggenberger (2003) to nonlinear models. Otsu (2006) also generalizes the findings to time series models and introduces an alternative Kleibergen type test. Guggenberger and Smith (2005) also have time series extension. Their research is carried independently of ours.

The differences between Guggenberger and Smith's (2005) paper and this one can be related to the contents of the proof methods. To see theoretical differences with the Guggenberger and Smith (2005) article, our consistency proof, which includes Lemma A.1 and Theorem 1, is new in the weak identification literature. 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 (2005) use in their paper benefit from the approach in Newey and Smith (2004).

Section 2 introduces the assumptions, the model, and consistency result. Section 3 considers tests that are robust to identification. Section 4 conducts simulations. Section 5 concludes. The appendix contains the consistency proof. For a more general case of generalized empirical likelihood estimators, the proofs can be found in Guggenberger and Smith (2005), and Otsu (2006). $\|.\|$ represents the Euclidean norm.

2 The Estimator

Suppose we are given the following moment conditions

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

where $\{x_t : t = 1, \cdots, T\}$ be the data vector of dimension "s" and $\psi(x_t, \theta)$ is an $r \times 1$ vector of observable real valued functions. x_t is iid. DGP is allowed to depend on T. This way of representing data is also used in Newey and Windmeijer (2007) in the case of many weak moments. Expectation in (1) is taken with respect to the distribution of x_t for sample size T, and we suppress dependence on T for notational convenience. The dependence on T is suppressed in these results for notational convenience. 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 $r \geq d$.

2.1 Model

As in Kitamura and Stutzer (1997) for all $\theta \in \Theta$

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

Next

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

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

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

where we set

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

Define the following matrices

$$\Omega(\theta) = E \psi_t(\theta) \psi_t(\theta)'.$$

$$\hat{\Omega}(\theta) = T^{-1} \sum_{t=1}^{T} \psi_t(\theta) \psi_t(\theta)'.$$

Let $\theta = (\alpha', \beta')'$, where $\alpha \in A, A \subset R^{d_1}, \beta \in B, B \subset R^{d_2}, d_1 + d_2 = d$. $\Theta = A \times B$ is a compact subset of R^d. A and B are also compact sets. In the following assumptions α are weakly identified whereas β are strongly identified. This can be seen in Assumptions 4 and 5. We need the following assumptions.

Assumptions:

1. $\Omega(\theta)$ is bounded on Θ. $\Omega(\theta)$ is nonsingular for all $\theta \in A \times B$

$$\sup_{\theta \in \Theta} \|\hat{\Omega}(\theta) - \Omega(\theta)\| = o_p(1),$$

$$\sup_{\theta \in A \times B} T^{-1} \sum_{t=1}^{T} \|\psi_t(\theta) \psi_t(\theta)'\| = O_p(1).$$

2. $E[\sup_{\theta \in \Theta} e^{\delta g' \psi_t(\theta)}] < \infty$ for all vectors g in a neighborhood of the origin, $\delta > 2$, and for all $1 \leq t \leq T$, $T \geq 1$.

3. i) $x_t$is iid.

ii)

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

for all $1 \leq t \leq T$, $T \geq 1$.

iii)

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

for all $\theta_1, \theta_2 \in \Theta$ and where $EB_t^\delta < \infty$, for some $\delta > 2$, and for all $1 \leq t \leq T$, $T \geq 1$.

4. $E\psi_t(\theta_0) = 0$ and for $\theta \neq \theta_0$ we have $E\psi_t(\theta) \neq 0$.

5. As $T \to \infty$, uniformly over $(\alpha, \beta) \in A \times B$

$$E\|\psi_t(\alpha, \beta) - \psi_t(\alpha_0, \beta)\|^2 \to 0.$$

Remarks. Assumptions 1 is used in the consistency proof, and is used in Guggenberger and Smith (2005). Assumption 2 is standard in this literature, as shown in Kitamura and Stutzer (1997). Assumption 2 is slightly more restrictive than the one in Kitamura and Stutzer (1997). Assumption 3 is used in deriving the uniform law of large numbers as Assumption B' in Stock and Wright (2000). Assumption 4 is uniqueness of the solution of θ_0. Assumption 5 is the weak identification assumption similar to the one used in Stock and Wright (2000). In this assumption, α are weakly identified, and β are identified. Note that Stock and Wright (2000) need the following for the weak identification assumption for the iid random variables:

$$E\psi_t(\alpha, \beta) = \frac{m_1(\theta)}{T^{1/2}} + m_2(\beta), \tag{5}$$

where equation (2.4) of Stock and Wright (2000) specifies, taking into account iid random variables

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

and $m_2(\beta) = 0$ only when $\beta_0 = 0$, and $m_1(\theta_0) = 0$. So

$$E[\psi_t(\alpha, \beta) - \psi_t(\alpha_0, \beta)] \to 0,$$

at rate $T^{1/2}$. Even though Assumption 5 is an assumption on second moments and hence slightly stronger than Stock and Wright (2000), unlike them it does not specify a rate of convergence to zero.

2.2 Consistency

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 Lemma A.1 in the Appendix and the proof of Theorem 1. The proof technique also differs from Guggenberger and Smith (2005) where they extend the proof of Newey and Smith (2004) to weakly identified parameters. We mainly focus on the behavior of the objective function directly to derive the consistency.

Theorem 1 . *Under Assumptions 1-5,*

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

The limit results are already established in Generalized Empirical Likelihood by Guggenberger and Smith (2005), and Otsu (2006).

3 Testing

The limits of estimators depends on nuisance parameters and these estimators are not consistent in weakly identified Generalizes Empirical Likelihood framework of Guggenberger and Smith (2005). 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).

Here we introduce a similar test in an exponential tilting framework with weak instruments. In order to understand the test statistic better we set the following notation

$$\Omega(\theta_0) = E\psi_t(\theta_0)\psi_t(\theta_0)'.$$

First we make the following Assumptions :

Assumption T.1.

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

Assumption T.2. *Assumption 1 holds at θ_0 and Assumption 4 holds .*

Next we provide the limit theory for tests. The proofs of Theorems 2 and 3 can be found in Guggenberger and Smith (2005), Otsu (2006).

Theorem 2. *Under Assumptions T.1, T.2, and with the following null hypothesis* $H_0 : \theta = \theta_0$,

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

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. One drawback of the Anderson-Rubin like test is it may have low power when the number of weak instruments are large.

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 (2005) test statistic for weakly identified GMM . We need the notation below before the following assumption. Denote

$$p_t(\theta_0) = (p_{1,t}(\theta_0)', \cdots, p_{d,t}(\theta_0)')',$$

where

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

for $i = 1, \cdots, d$. $p_{i,t}(\theta_0)$ is $r \times 1$ vector, and $p_t(\theta_0)$ is $dr \times 1$ vector. Next define

$$\bar{p}_T(\theta_0) = \frac{1}{T} \sum_{t=1}^{T} [p_t(\theta_0) - Ep_t(\theta_0)],$$

and

$$\bar{\Psi}_T(\theta_0) = T^{-1} \sum_{t=1}^{T} \psi_t(\theta_0). \tag{6}$$

The following assumption is first used in Kleibergen (2005).

Assumption T.3. *The joint limiting behavior of the averages* $\bar{\Psi}_T(\theta_0)$ *and* $\bar{p}_T(\theta_0)$ *satisfy the following Central Limit Theorem*

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

where $\Psi(\theta_0)$ *is* $r \times 1$, *and* $\Psi_p(\theta_0)$ *is* $dr \times 1$ *vector. The following* $V(\theta_0)$ *is* $(r + dr) \times (r + dr)$ *positive semidefinite symmetric matrix.*

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

where dimensions of the sub matrices are $\Omega(\theta_0) : r \times r$, $\Sigma(q, \theta_0) : dr \times r$, $\Sigma(q, q) : dr \times dr$ *and* $\Sigma(q, \theta_0) = \Sigma(\theta_0, q)'$. *Also*

$$\Sigma(\theta_0, q) = E[(\psi_t(\theta_0))(p_t(\theta_0) - Ep_t(\theta_0))'],$$

$$\Sigma(q, q) = E[(p_t(\theta_0) - Ep_t(\theta_0))(p_t(\theta_0) - Ep_t(\theta_0))'].$$

Assumption T.4.

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

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

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

Assumption T.3 assumes the existence of a simple Central Limit Theorem. These assumptions are standard in this literature, as seen in Kitamura and Stutzer (1997), Kleibergen (2005), and Stock and Wright (2000). Detailed discussion about sufficient conditions and positive semidefiniteness of $V(\theta_0)$ can be found in Kleibergen (2005). In Assumption T.4, an asymptotically equivalent estimator for $\Omega(\theta_0)$ is $\hat{\Omega}(\theta_0) = \frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)\psi_t(\theta_0)'$. But we think $\tilde{\Omega}(\theta_0)$ is better in finite samples.

As in Kleibergen (2005), we benefit from the first order condition in exponential empirical likelihood

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

We base this test statistic on an asymptotically equivalent form of the first order condition.

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

$$\frac{\partial \hat{Q}_T(\theta, \hat{\gamma}(\theta))}{\partial \theta'}|_{\theta_0} = \hat{\gamma}(\theta_0)'\bar{D}_T(\theta_0), \tag{7}$$

where

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

$\bar{D}_T(\theta_0)$ is of dimension $r \times d$. Note that (7) is a simplified version of the actual first order condition. The algebraic simplifications to reach (7) are shown in the proof of Theorem 2 in Kitamura and Stutzer (1997).

The following Theorem extends the GMM K-statistic in Kleibergen (2005) to exponential tilting estimators. Note that Guggenberger and Smith (2005) also consider the K-statistic in generalized empirical likelihood models. The limit in the following Theorem 5 is the same as in Kleibergen (2005).

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

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

where

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

and $\bar{\Psi}_T(\theta_0)$ is described in (6).

We can easily show that, the K-statistic in our case is asymptotically equivalent to the K-statistic in Kleibergen (2005) for CUE. The main difference between the K-test in Kleibergen (2005) 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 3, the subtests can be developed easily simply by following sections 3.2 of Kleibergen (2005).

The K-statistic in the continuously updating estimator case of Kleibergen (2005) takes the value of zero when the GMM objective function is at its minimum, maximum, and inflection points. This results in spuriously low power of K-test statistic in CUE. So in section 5 of Kleibergen (2005), a pretest of moment conditions before testing $H_0 : \theta = \theta_0$ is introduced. Note that since the K-test statistic here does not depend on an exact first order condition in (7), it does not take the value zero when the moment restrictions are invalid. If we had instead built our test using (7) at $\hat{\theta}$, this test could have taken the value of zero at its maximum point, and inflection points at objective function. 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 than an alternative K-test which uses (7). So there is no need to introduce such a pretest of the moment conditions in our case. By inverting the Anderson-Rubin like test statistic and the K-test statistic we can have confidence intervals for θ.

4 Simulation

In this section we analyze the size and power of the Anderson-Rubin like and Kleibergen test statistics . We compare Continuous Updating (CUE), Empirical Likelihood (EL) and Exponential Tilting (ET) based tests. Here we benefit from the setup used by Guggenberger and Smith (2005). Our difference will be the error distributions that are used. We use two designs, one involving GARCH (1,1) structural errors, and the other one involving a contaminated structural error distribution. First, our structural and reduced form equations are

$$y = Y\theta_0 + u,$$

$$Y = Z\Pi + V.$$

There will be only one single endogenous variable Y, and no included exogenous variables.

$$Z \sim N(0, I_r \otimes I_n),$$

where r is the number of instruments. The number of instruments vary between 1 and 5. We vary $\Pi = \{0.02, 0.1, 0.4\}$ for $r = 1$, and Π takes the same value for each cell in $r = 5$ case. $r = 1$ case represents just identified system with one instrument, and $r = 5$ case represents the system with many instruments. In this respect, our $r = 5$ setup is different from Guggenberger and Smith (2005), since they use 1 weak and 4 irrelevant instruments ($\Pi = 0$ for these irrelevant ones). For the size exercise the true value of $\theta_0 = 0$. We test $\theta_0 = 0$ versus $\theta_0 \neq 0$ in the simulation.

For both designs $(u, V)'$ is distributed as $N(0, \Sigma \otimes I_n)$, where

$$\Sigma = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix}.$$

The first setup has GARCH (1,1) structural errors. Then instead of u_t in structural errors, we use ϵ_t which follows GARCH (1,1), but u_t and ϵ_t is related in the following way:

$$\epsilon_t = h_t^{1/2} u_t,$$

$$h_t = \Xi + \delta_1 \epsilon_{t-1}^2 + \delta_2 h_{t-1},$$

where $t = 1, \cdots, T$, and $\Xi = 0.1, \delta_1 = 0.4, \delta_2 = 0.3$. Then in the second setup, we use the following contaminated error ϵ_t instead of u_t in the structural equation:

$$\epsilon_t = .9u_t + .1\omega_t,$$

where ω_t is independent of u_t, and ω_t is distributed as $N(0, 16)$.

The first setup examines the size and power of the tests under CUE, EL, ET frameworks with respect to GARCH (1,1) structural errors. The second setup analyzes the small sample properties of these tests under CUE, EL, ET when we have contaminated structural errors. These designs extend the former designs in the weak identification literature to specific heavy tailed errors and to contaminated design. We use Anderson-Rubin like test in CUE, EL, ET techniques. To ensure a fair comparison between them we provide the Anderson-Rubin like test in Generalized Empirical Likelihood Estimation framework provided by Guggenberger and Smith (2005):

$$AR = 2\sum_{t=1}^{T} \rho(\gamma' \psi_t(\theta))/T - 2\rho(0),$$

where for CUE, $\rho(v) = (1 + v)^2/2$, for ET $\rho(v) = exp(v)$, for EL $\rho(v) = -ln(1 - v)$. Note that AR test in ET is different than from the one in Theorem 2. This new asymptotically equivalent test invites a fair comparison between CUE and EL. However, we also have simulated the test in Theorem 2, this test is denoted as LET in tables. Since we maximize the objective function with respect to θ we use positive sign for CUE, ET and negative sign for EL.

For Kleibergen type of test we use Theorem 3, but $\bar{D}_T(\theta_0)$ is obtained differently in EL and CUE framework, where

$$\bar{D}_T(\theta_0) = \sum_{t=1}^{T} \rho_1(\gamma(\theta_0)'\psi_t(\theta_0))/T,$$

where $\rho_1(.)$ is the partial derivative of $\rho(.)$ above.

4.1 Size

The results for GARCH (1,1) case and contaminated errors are very similar. We discuss GARCH (1,1) and contaminated design results which are on Table 1 for $\Pi = .1$ and Anderson-Rubin type of tests. We see that with small number of observations at GARCH (1,1) case , size of the AR test is 18.5 for EL, 13.7 for ET, 16.1 for log ET (LET) and 8.3 for CUE at 5% level with 5 instruments. Generally among Anderson-Rubin type of tests, CUE based one is the best and ET based one also has very good size. EL based test has high size distortions in this case. Among Kleibergen type of tests in Table 2 with $\Pi = .1$, ET based one has excellent size with 5-7% depending on sample size and number of instruments. CUE based test has 7-12% size and EL based one has 5-9% size. So overall from Tables 1 and 2, ET based Kleibergen type test has the best size. Tables 5-6, 9-10 also show size of the tests at $\Pi = .02$, $\Pi = .4$ respectively. These tables demonstrate that the size of the tests does not change much with Π.

4.2 Power

The alternatives are $\theta_0 = \{-1, -0.8, -0.6, 0.6, 0.8, 1\}$. We analyze the case of very weak instruments ($\Pi = .02$), weak instruments ($\Pi = .1$), and strong instruments ($\Pi = .4$). From Tables 3-4, 7-8 for the cases of very weak/weak instruments we see that none of the tests have good power. Only in the case of many weak instruments we see that CUE based K-test does better than others. In the case of strong instruments all the tests have good power. This can be observed from Tables 11-12. All tests have more power in GARCH (1,1) design compared with contaminated errors. The power is size-adjusted and critical values are available from the author on demand. Sample size in the tables is 100.

5 Conclusion

This paper investigates asymptotics of exponential tilting estimators in the case of weak identification. These are very different from the asymptotically normal ones. We also derive test statistics

Table 1: Size at 5% level, ANDERSON-RUBIN TESTS, $\Pi = .1$

	GARCH (1,1)				CONTAMINATED DESIGN			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$
EL	7.2	18.5	6.2	11.3	6.9	18.4	6.1	10.3
ET	6.7	13.7	6.1	9.7	6.5	13.5	5.6	9.0
LET	7.1	16.1	6.4	10.1	6.8	16.0	5.8	9.9
CUE	5.5	8.3	5.9	6.6	5.6	8.3	5.4	8.5

Note: The test statistics are compared to 5% critical values of the limits in Theorem 2 and Theorem 3. For Anderson-Rubin test for $r = 1, \chi^2_1 = 3.84$*; for* $r = 5$*,* $\chi^2_5 = 11.06$*. We conduct 10000 trials.*

Table 2: Size at 5% level, KLEIBERGEN TESTS, $\Pi = .1$

	GARCH (1,1)				CONTAMINATED DESIGN			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$
EL	5.3	8.4	5.1	6.7	5.3	8.6	5.7	6.7
ET	5.5	7.0	5.0	5.9	5.9	6.9	5.5	5.6
CUE	5.7	12.2	4.9	8.1	5.7	9.7	5.3	7.0

Note: The test statistics are compared to 5% critical values of the limits in Theorem 2 and Theorem 3. For the Kleibergen test for all cases $\chi^2_1 = 3.84$*. We conduct 10000 trials.*

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 many weak instruments framework.

Table 3: Size Adjusted Power, ANDERSON-RUBIN TESTS, $\Pi = .1$

	GARCH (1,1), EL						CONTAMINATED DESIGN, EL					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	13.9	13.7	12.9	9.8	10.7	11.6	11.3	10.6	8.9	8.8	8.6	10.0
$r = 5$	29.6	24.5	22.1	15.0	17.6	19.1	19.1	16.6	13.6	7.4	9.8	14.3
	GARCH (1,1), ET						CONTAMINATED DESIGN, ET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	15.7	15.1	12.8	9.4	10.1	12.1	13.5	10.4	8.9	9.1	9.1	10.0
$r = 5$	28.2	25.9	21.6	15.6	17.0	17.8	23.1	16.4	12.9	11.1	13.6	14.9
	GARCH (1,1), LET						CONTAMINATED DESIGN, LET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	16.0	13.4	14.5	10.6	9.9	12.3	14.5	10.2	7.1	8.4	7.7	10.0
$r = 5$	28.2	28.3	22.6	14.2	16.8	20.6	21.1	17.5	12.3	10.7	13.2	13.5
	GARCH (1,1), CUE						CONTAMINATED DESIGN, CUE					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	14.1	13.2	11.9	8.9	10.9	12.0	14.8	12.7	12.5	10.1	7.5	9.6
$r = 5$	30.3	24.1	21.1	11.8	16.3	16.7	25.1	15.7	12.7	11.3	7.2	9.8

Note: The test statistics are compared to finite sample critical values that are obtained by running the same size program in Table 1 . These can be obtained from the author on request. We conduct 1000 trials. $T = 100$.

Table 4: Size Adjusted Power, KLEIBERGEN TESTS, $\Pi = .1$

	GARCH (1,1), EL						CONTAMINATED DESIGN, EL					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	15.9	15.1	13.0	11.5	11.4	11.7	13.6	10.8	9.4	5.9	6.9	10.0
$r = 5$	16.1	17.0	17.4	17.1	16.4	15.8	20.0	16.0	12.8	8.8	13.0	15.3
	GARCH (1,1), ET						CONTAMINATED DESIGN, ET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	16.0	14.8	11.3	10.7	12.0	12.0	12.8	10.4	8.4	6.9	8.7	11.0
$r = 5$	16.3	17.2	15.4	15.0	14.2	17.6	14.6	16.0	11.8	10.4	11.7	15.6
	GARCH (1,1), CUE						CONTAMINATED DESIGN, CUE					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	15.6	17.0	14.8	9.5	12.1	13.8	13.7	11.4	8.3	8.5	8.4	9.3
$r = 5$	36.1	33.4	25.8	29.4	35.0	38.3	29.6	26.2	14.3	19.0	23.7	31.8

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 conduct 1000 trials. $T = 100$.

Table 5: Size at 5% level, ANDERSON-RUBIN TESTS, $\Pi = .02$

	GARCH (1,1)				CONTAMINATED DESIGN			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$
EL	7.1	19.0	6.0	11.5	6.6	17.5	5.5	10.5
ET	6.8	14.3	6.3	9.6	6.5	13.3	5.5	8.8
LET	6.8	16.4	6.1	10.8	6.8	15.1	5.6	10.4
CUE	5.4	7.9	5.5	5.9	6.0	7.8	5.6	6.4

Note: The test statistics are compared to 5% critical values of the limits in Theorem 2 and Theorem 3. For Anderson-Rubin test for $r = 1, \chi^2_1 = 3.84$*; for* $r = 5$*,* $\chi^2_5 = 11.06$*. We conduct 10000 trials.*

Table 6: Size at 5% level, KLEIBERGEN TESTS, $\Pi = .02$

	GARCH (1,1)				CONTAMINATED DESIGN			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$
EL	5.7	8.8	5.1	6.8	5.9	8.9	5.2	7.0
ET	5.8	7.2	5.5	5.8	5.6	7.0	5.6	5.8
CUE	5.4	10.5	5.3	7.9	5.8	8.6	5.4	6.9

Note: The test statistics are compared to 5% critical values of the limits in Theorem 2 and Theorem 3. For Anderson-Rubin test for $r = 1, \chi^2_1 = 3.84$*; for* $r = 5$*,* $\chi^2_5 = 11.06$*. We conduct 10000 trials.*

Table 7: Size Adjusted Power, ANDERSON-RUBIN TESTS, $\Pi = .02$

	GARCH (1,1), EL						CONTAMINATED DESIGN, EL					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	5.6	6.7	5.2	6.7	4.6	4.4	7.1	5.8	5.5	5.9	6.6	5.2
$r = 5$	4.6	4.8	4.6	4.9	4.3	4.2	6.3	6.6	6.7	5.3	6.3	6.2
	GARCH (1,1), ET						CONTAMINATED DESIGN, ET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	4.6	5.3	4.9	5.1	6.0	5.4	5.5	4.5	6.6	4.4	5.7	4.8
$r = 5$	6.2	5.2	5.2	5.7	4.7	5.1	3.9	5.3	6.5	6.1	6.5	6.1
	GARCH (1,1), LET						CONTAMINATED DESIGN, LET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	6.1	5.1	4.7	6.0	5.1	4.0	5.3	6.6	6.3	5.3	5.3	7.0
$r = 5$	5.6	5.8	5.7	4.1	5.3	4.4	5.5	5.1	5.9	4.2	5.8	5.1
	GARCH (1,1), CUE						CONTAMINATED DESIGN, CUE					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	6.0	5.8	4.8	5.6	4.4	5.0	5.3	6.4	4.5	5.6	4.2	4.8
$r = 5$	6.5	5.4	3.9	5.3	4.7	6.2	4.6	4.6	4.1	6.3	5.2	4.5

Note: The test statistics are compared to finite sample critical values that are obtained by running the same size program in Table 5 . These can be obtained from the author on request. We conduct 1000 trials. $T = 100$.

Table 8: Size Adjusted Power, KLEIBERGEN TESTS, $\Pi = .02$

	GARCH (1,1), EL						CONTAMINATED DESIGN, EL					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	6.1	5.2	5.3	3.7	5.1	5.3	5.3	4.5	6.3	4.1	3.4	5.1
$r = 5$	5.2	5.4	4.9	4.1	5.1	4.7	5.0	5.4	3.7	5.2	4.6	4.5
	GARCH (1,1), ET						CONTAMINATED DESIGN, ET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	6.8	6.4	6.5	4.2	4.8	5.5	5.0	4.8	4.5	5.8	4.5	5.5
$r = 5$	5.9	5.7	4.9	5.5	5.5	5.2	3.6	6.6	5.6	5.3	4.2	5.7
	GARCH (1,1), CUE						CONTAMINATED DESIGN, CUE					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	5.4	7.1	4.2	6.2	4.9	6.4	5.8	4.5	4.5	6.1	5.4	5.5
$r = 5$	4.8	4.0	4.2	5.2	6.4	6.3	4.2	4.3	5.0	6.7	5.4	7.5

Note: The test statistics are compared to finite sample critical values that are obtained by running the same size program in Table 6 . These can be obtained from the author on request. We conduct 1000 trials. $T = 100$.

Table 9: Size at 5% level, ANDERSON-RUBIN TESTS, $\Pi = .4$

	GARCH (1,1)				CONTAMINATED DESIGN			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$
EL	6.8	19.7	6.2	11.6	6.5	17.3	5.9	10.2
ET	6.8	14.5	5.9	9.3	6.1	13.9	6.2	8.9
LET	6.5	16.1	5.7	10.5	6.7	15.6	5.9	10.1
CUE	5.2	8.2	5.3	6.6	5.7	8.2	5.3	6.2

Note: The test statistics are compared to 5% critical values of the limits in Theorem 2 and Theorem 3. For Anderson-Rubin test for $r = 1, \chi_1^2 = 3.84$*; for* $r = 5$*,* $\chi_5^2 = 11.06$*. We conduct 10000 trials.*

Table 10: Size at 5% level, KLEIBERGEN TESTS, $\Pi = .4$

	GARCH (1,1)				CONTAMINATED DESIGN			
	$T = 50$		$T = 100$		$T = 50$		$T = 100$	
	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$	$r = 1$	$r = 5$
EL	5.3	7.4	5.5	5.9	5.7	7.4	5.3	6.1
ET	5.3	6.1	5.5	5.5	5.7	6.2	5.6	5.1
CUE	5.8	16.2	5.2	12.2	6.1	15.6	5.2	11.5

Note: The test statistics are compared to 5% critical values of the limits in Theorem 2 and Theorem 3. For Anderson-Rubin test for $r = 1, \chi_1^2 = 3.84$*; for* $r = 5$*,* $\chi_5^2 = 11.06$*. We conduct 10000 trials.*

Table 11: Size Adjusted Power, ANDERSON-RUBIN TESTS, $\Pi = .4$

	GARCH (1,1), EL						CONTAMINATED DESIGN, EL					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	95.5	93.7	87.8	70.3	78.7	84.8	87.1	76.4	60.4	48.4	56.7	71.8
$r = 5$	100.0	100.0	100.0	98.4	99.7	99.9	99.9	99.7	97.2	85.5	97.1	98.7
	GARCH (1,1), ET						CONTAMINATED DESIGN, ET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	95.9	93.6	87.7	72.0	83.5	84.8	87.4	77.2	62.3	44.3	59.2	67.2
$r = 5$	99.3	99.9	99.0	98.6	98.7	95.1	97.4	99.5	96.7	84.7	89.7	82.7
	GARCH (1,1), LET						CONTAMINATED DESIGN, LET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	95.8	93.7	88.3	73.7	83.5	86.4	88.6	80.2	61.8	46.9	58.8	68.4
$r = 5$	99.1	100.0	99.5	98.5	98.4	95.0	97.2	97.1	97.4	87.1	90.4	79.3
	GARCH (1,1), CUE						CONTAMINATED DESIGN, CUE					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	95.3	94.1	88.6	71.9	80.9	85.7	86.7	78.2	62.5	46.7	59.9	71.8
$r = 5$	100.0	99.9	99.9	98.5	99.7	99.8	100.0	99.7	96.3	86.0	94.4	98.4

Note: The test statistics are compared to finite sample critical values that are obtained by running the same size program in Table 9 . These can be obtained from the author on request. We conduct 1000 trials. $T = 100$.

Table 12: Size Adjusted Power, KLEIBERGEN TESTS, $\Pi = .4$

	GARCH (1,1), EL						CONTAMINATED DESIGN, EL					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	96.1	94.6	89.6	74.2	81.0	84.1	82.7	80.4	61.6	47.1	59.0	68.6
$r = 5$	65.5	72.0	82.9	96.1	95.1	91.6	93.3	96.3	97.6	93.6	97.7	97.6
	GARCH (1,1), ET						CONTAMINATED DESIGN, ET					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	96.4	93.6	89.0	69.2	80.8	86.2	86.8	79.1	62.4	46.5	58.2	66.8
$r = 5$	63.0	73.0	87.3	98.0	95.0	92.7	93.1	95.4	96.6	92.0	96.7	97.7
	GARCH (1,1), CUE						CONTAMINATED DESIGN, CUE					
θ_0	-1	-0.8	-0.6	0.6	0.8	1	-1	-0.8	-0.6	0.6	0.8	1
$r = 1$	95.4	93.3	86.5	71.0	79.3	87.4	87.4	80.2	64.2	45.8	61.5	67.9
$r = 5$	100.0	99.7	100.0	99.8	99.9	100.0	100.0	99.9	98.0	95.6	99.4	100.0

Note: The test statistics are compared to finite sample critical values that are obtained by running the same size program in Table 10 . These can be obtained from the author on request. We conduct 1000 trials. $T = 100$.

APPENDIX

We need the following lemma for the consistency proof.

Lemma A.1. *Under Assumptions 2,3 and 5, uniformly in* $(\alpha, \beta) \in A \times B$,

$$\frac{1}{T}\sum_{t=1}^{T}[e^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)} - Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)}] \xrightarrow{p} 0.$$

Proof of Lemma A.1. First we need to prove the following uniform law of large numbers

$$T^{-1}\sum_{t=1}^{T}[e^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)} - Ee^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)}] \xrightarrow{p} 0. \tag{8}$$

At $\theta = (\alpha', \beta')' \in \Theta$, note that Jacobian of the first order condition for (2) is nonsingular under Assumption 1. So $\gamma(\theta)$ is continuous in its argument. Same point is observed by Kitamura and Stutzer (1997) in their proof of Theorem 1. Note that $\gamma(\theta)$ is continuous over Θ which is a compact set. Then by Theorem 2.19 of Davidson (1994) $\gamma(\theta)$ is bounded and uniformly continuous.

Furthermore by proof of Proposition 17 on p.47 of Royden (1988), and $\gamma(\theta)$ being not a sequence by (2)

$$|\gamma(\theta)| \leq \sup_{\theta\in\Theta}|\gamma(\theta)| < \infty. \tag{9}$$

It is clear that bounds do not depend on T.

Next by (9) for all $\theta_1, \theta_2 \in \Theta$, and by Assumption 3

$$|\gamma(\theta_2)'\psi_t(\theta_2) - \gamma(\theta_1)'\psi_t(\theta_2)| \leq [\sup_{\theta}|\gamma(\theta)|]B_t|\theta_2 - \theta_1|. \tag{10}$$

Basically (10) is the Lipschitz condition for $\gamma(\theta)'\psi_t(\theta)$. Then Assumption 2 with (10) provides the Lipschitz continuity of $e^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)}$. By Assumption 3i, and using Assumption 2 with Lipschitz continuity of $e^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)}$, we have the uniform law of large numbers in (8). To clarify this point further, by these conditions as in Theorems 1 and 2 of Andrews (1994), we can obtain weak convergence of empirical process based on the function $e^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)}$ which implies the uniform law of large number. This point can also be seen on p.673 of Guggenberger and Smith (2005).

Next in the proof we have to show that as $T \to \infty$, uniformly over $(\alpha, \beta) \in A \times B$,

$$Ee^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)} - Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)} \to 0. \tag{11}$$

By mean-value expansion there exists vectors such as $\bar{\gamma}, \bar{\psi}_t$

$$\begin{aligned} Ee^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)} &- Ee^{\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)} \\ &= E[\gamma(\alpha,\beta)'\psi_t(\alpha,\beta) - \gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)]e^{\bar{\gamma}'\bar{\psi}_t}, \end{aligned} \tag{12}$$

where without losing any generality, $\bar{\gamma}'\bar{\psi}_t \in (\gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta), \gamma(\alpha,\beta)'\psi_t(\alpha,\beta))$.

Then by (9), Cauchy-Schwartz inequality with Assumptions 2 and 5

$$
\begin{aligned}
E[\gamma(\alpha,\beta)'\psi_t(\alpha,\beta) - \gamma(\alpha_0,\beta)'\psi_t(\alpha_0,\beta)]e^{\bar{\gamma}'\bar{\psi}_t} &\le [\sup_{(\alpha,\beta)\in A\times B} |\gamma(\alpha,\beta)|]E(\psi_t(\alpha,\beta) - \psi_t(\alpha_0,\beta))e^{\bar{\gamma}'\bar{\psi}_t} \\
&\le [\sup_{(\alpha,\beta)\in A\times B} |\gamma(\alpha,\beta)|][E\|\psi_t(\alpha,\beta) - \psi_t(\alpha_0,\beta)\|^2]^{1/2} \\
&\times [Ee^{2\bar{\gamma}'\bar{\psi}_t}]^{1/2} \\
&\to 0. \qquad (13)
\end{aligned}
$$

So (13) proves (11). Then we combine (11) with (8) to have the desired result. **Q.E.D**.

The proof of (8) can be done with slightly different condition than in Assumption 3iii. Change the Assumption 3iii to the following

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

holds almost surely, for all $\theta_1, \theta_2 \in \Theta$, for all $1 \le t \le T$, and $B_t = O_p(1)$. This is a more direct proof since this is sufficient condition for stochastic equicontinuity in Theorem 21.10 of Davidson (1994) of uniform law of large numbers. Assumption 2 and (10) then provide uniform law of large number in (8).

Proof of Theorem 1. The first part of the proof proceeds very similar to equations (13)-(14) of Kitamura and Stutzer (1997).

First, Assumption 4 implies that there is a unique saddle point at $(\alpha_0, \beta_0, \gamma(\alpha_0, \beta_0))$ of the function $M(\alpha, \beta, \gamma(\alpha, \beta)) = Ee^{\gamma(\alpha,\beta)'\psi_t(\alpha,\beta)}$. The value of the function at the saddle point $M(\alpha_0, \beta_0, \gamma(\alpha_0, \beta_0)) = 1$. Assumption 4 also implies, at $\alpha = \alpha_0, \beta \neq \beta_0$, by equation (13) of Kitamura and Stutzer (1997)

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

Let $\Gamma(\beta, \delta)$ denote an open sphere with center β and radius δ. Next proceeding in a similar way as in p.869 of Kitamura and Stutzer (1997), using Assumptions 1-3 and analyzing the parameter space $\Theta - \Gamma(\beta_0, \delta)$ we obtain

$$E[\sup_{\beta'\in\Theta-\Gamma(\beta_0,\delta)} e^{\gamma(\alpha_0,\beta')'\psi_t(\alpha_0,\beta')}] = 1 - 2h, \qquad (14)$$

where h is a small positive number. Use Lemma A.1 to have

$$P[\sup_{(\alpha,\beta')\in\Theta-\Gamma(\beta_0,\delta)} \frac{1}{T}\sum_{t=1}^{T} \left(e^{\gamma(\alpha,\beta')'\psi_t(\alpha,\beta')} - Ee^{\gamma(\alpha_0,\beta')\psi_t(\alpha_0,\beta)}\right) > h] < \epsilon/2. \qquad (15)$$

Consider (14)-(15) to have, for small $\epsilon > 0$

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

By (16)-(17) of Kitamura and Stutzer (1997), noting that $\hat{\gamma}(.)$ is defined in (4), and $\gamma(.)$ is defined in (2),

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

For large T,

$$P[\sup_{(\alpha,\beta')\in\Theta-\Gamma(\beta_0,\delta)} \frac{1}{T}\sum_{t=1}^{T} e^{\hat{\gamma}(\alpha,\beta')'\psi_t(\alpha,\beta')} > 1-h] < \epsilon/2. \tag{16}$$

But from Lemma A.1 and equation (15) it is clear that in the large samples α is not identified and only the consistency of β is relevant.

First, uniformly over α (including α_0) under Assumption 3

$$\frac{1}{T}\sum_{t=1}^{T} \psi_t(\alpha,\beta_0) = O_p(1). \tag{17}$$

This can also be seen by Remark on p.673 of Guggenberger and Smith (2005).

Next under Assumptions 1-3 by Lemmata 8-9 of Guggenberger and Smith (2005) the following exists uniformly with probability approaching one

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

Then via Lemma 9 of Guggenberger and Smith (2005) uniformly over $\alpha \in A$ (including α_0)

$$\hat{\gamma}(\alpha,\beta_0) \xrightarrow{p} 0.$$

By convexity, for all $\alpha \in A$,

$$e^{\hat{\gamma}(\alpha,\beta_0)'\sum_{t=1}^{T}\psi_t(\alpha,\beta_0)/T} \leq \frac{1}{T}\sum_{t=1}^{T} e^{\hat{\gamma}(\alpha,\beta_0)'\psi_t(\alpha,\beta_0)} \leq 1.$$

Note that the last equation is true at $\alpha = \alpha_0$ as well.

Combine the last three equations to have uniformly over $\alpha \in A$

$$e^{\hat{\gamma}(\alpha,\beta_0)\sum_{t=1}^{T}\psi_t(\alpha,\beta_0)/T} \xrightarrow{p} 1,$$

where A includes α_0. The last two equations then show

$$P[\sup_{\alpha\in A} \frac{1}{T}\sum_{t=1}^{T} e^{\hat{\gamma}(\alpha,\beta_0)'\psi_t(\alpha,\beta_0)} < 1-h/2] < \epsilon/2. \tag{18}$$

Then Lemma A.1 and (15)-(18) imply consistency of $\hat{\beta}$. The main difference with the consistency proof for all well identified parameters in Kitamura and Stutzer (1997) is Lemma A.1, and equations (15)-(18). These show that weakly identified parameter vector α is not consistent. **Q.E.D**.

REFERENCES

Anderson, T.W. & 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.

Caner, M. (2007) Boundedly Pivotal Structural Change Tests in Continuous Updating GMM with strong, weak identification and completely unidentified cases. *Journal of Econometrics,* 137, 28-67.

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

Guggenberger, P. (2003) Econometric essays on generalized empirical likelihood, long memory time series and volatility. Ph. D Thesis, Yale University.

Guggenberger, P. & R.J. Smith (2005) Generalized Empirical Likelihood Tests under Partial, Weak and Strong Identification. *Econometric Theory,* 21, 667-709.

Imbens, G.W., R.H. Spady, & P. Johnson (1998) Information Theoretic Approaches To Inference In Moment Condition Models. *Econometrica,* 66, 333-357.

Kitamura, Y. & M. Stutzer (1997) An Information-Theoretic Alternative to Generalized Method of Moments Estimation. *Econometrica,* 65, 861-875.

Kleibergen, F. (2005) Testing Parameters in GMM Without Assuming That They Are Identified. *Econometrica*, 73, 1103-1123.

Newey, W.K. & R.J. Smith (2004) Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators. *Econometrica*, 72, 219-255.

Newey, W.K. & F. Windmeijer (2007) GMM with Many Weak Moment Conditions. Working Paper. Department of Economics, MIT.

Otsu, T. (2006) Generalized empirical likelihood inference for nonlinear and time series models under weak identification. *Econometric Theory*, 22, 513-529.

Royden, H.L. (1988) *Real Analysis*, Prentice Hall Press.

Stock, J.H., & J.H. Wright (2000) GMM With Weak Identification. *Econometrica,* 68, 1055-1096.

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.