This article aims at achieving two distinct goals. The first is to extend the existing LM test of overdispersion to the situation where the alternative hypothesis is characterized by the correlated random effects model. We obtain a result that the test against the random effects model has a certain max-min type optimality property. We will call such a test the LM test of overdispersion. The second goal of the article is to draw a connection between panel data analysis and the analysis of multiplicity of equilibrium in games. Because such multiplicity can be viewed as a particular form of neglected heterogeneity, we propose an intuitive specification test for a class of two-step game estimators.

KEY WORDS: Correlated random effects; LM test of overdispersion; Multiplicity of equilibrium.

1. INTRODUCTION

Econometric model specification is usually tested by exploiting one of the three principles, including (i) the Wald test, which is based on the asymptotic distribution of parameter estimates; (ii) the likelihood ratio (LR) test; and (iii) the Lagrange multiplier (LM) procedure, which is based on the derivative of the log-likelihood (score) imposing the hypothesis. (The relationship among the three tests was reviewed by, e.g., Engle (1984).) The LM procedure seems to dominate the other two in terms of convenience when the objective of the test is to detect neglected heterogeneity in a given econometric model. The LM test statistic is solely based on the parameter estimates under the null hypothesis that there is no neglected heterogeneity and, as such, it eliminates the burden of specifying and estimating the model under the alternative hypothesis that there is some neglected heterogeneity.

The possibility of using the LM test as a way of detecting individual heterogeneity was discussed by Breusch and Pagan (1980), Chesher (1984), Lee and Chesher (1986), and Hahn, Newey, and Smith (2014), among others. These tests can be viewed as a version of White’s (1982) information matrix test, as discussed by Chesher (1984). Alternatively, they can be understood to be a test of overdispersion as in Cox (1983), which is our interpretation in this article. Implicit in these tests is the assumption that the neglected heterogeneity (under the alternative hypothesis) is more or less independent of the randomness that generates the main model, although the assumption is made more explicit in Hahn, Newey, and Smith (2014, Lemmas 3.1 and 3.2). In the panel data context, this assumption amounts to the random effects specification, as developed in Balestra and Nerlove (1966) or Maddala (1971). In other words, in panel data applications we may roughly say that these tests are designed to detect random effects, but not fixed effects (or equivalently, correlated random effects). It is therefore of interest to revisit the LM test and analyze the test of neglected heterogeneity under more general conditions.

This article aims at achieving two distinct goals. The first is to extend the existing LM test to the situation where the alternative hypothesis is characterized by the fixed effects model, where we adopt the convention of equating the fixed effects model with the correlated random effects model (see, e.g., Chamberlain 1984). We begin by analyzing a model where the (local) alternative hypothesis is characterized by an arbitrary correlated random effects model, where the mean independence assumption is violated. We obtain a result that the test to detect such an arbitrary correlated random effects would take the form of the test of conditional moment restriction. We then go on to analyze a more restricted version of correlated random effects, where the effects are assumed to be mean independent of the covariates. We then go on to analyze a more restricted version of correlated random effects, where the effects are assumed to be mean independent of the covariates. For the latter case, we obtain an interesting result that the test against the random effects model, where the conditional variance of the effects does not depend on the covariates, has a certain max-min type optimality property. We will call such a test the LM test of overdispersion. In addition, while we show the LM test of overdispersion has no power in linear models where the mean independence assumption is violated, we show the test has some robustness properties against such violations in nonlinear models. As an example, we show the test has power...
against the arbitrary correlated random effects alternative in the case of the logit model commonly used in applications.

These results suggest that the most natural applications for the LM test of overdispersion are nonlinear models where the mean independence assumption is expected to hold. We find such an application in the context of two-step semiparametric estimation of Bayesian game models commonly applied to the analysis of environments with social or strategic interactions among agents. This leads to the second goal of the article, which is to draw a connection between panel data analysis and the analysis of multiplicity of equilibrium in games. Such multiplicity can be viewed as a particular form of neglected heterogeneity, and we propose an intuitive specification test for a class of two-step game estimators. Moreover, the canonical specification of these games is a system of logit discrete choice models, to which we show the test is robust against more general forms of neglected heterogeneity that may arise from elsewhere in the model.

Two-step estimators of Bayesian game models exploit the two requirements of Bayes–Nash equilibrium. (See, e.g., Aguirregabiria and Mira 2002; Bajari et al. 2011.) First, equilibrium requires that players’ strategies are optimal given their beliefs about the distribution of opponent play and, second, that these beliefs are consistent with the true equilibrium distribution, conditional on any common knowledge determinants of player payoffs. Under some assumptions about the nature of unobserved heterogeneity, the latter assumption allows the econometrician to, in the first step, nonparametrically estimate player beliefs directly from the data by estimating conditional equilibrium choice probabilities. Treating these estimates as data, the problem of estimating the model of interdependent actions reduces to one of estimating a single agent choice model in the second step. This two-step procedure allows the model to be estimated without ever having to solve the model, an often computationally expensive task. (This insight has been particularly important in the literature on estimating dynamic games (Aguirregabiria and Mira 2007; Bajari, Benkard, and Levin 2007; Pakes, Ostrovsky, and Berry 2007; Pesendorfer and Schmidt-Dengler 2008), where calculating equilibrium is so computationally expensive that “full solution” approaches to estimation are generally infeasible.)

The first-stage estimates conditional choice probabilities generally require pooling across a large number of observed game outcomes, which necessitates restrictive assumptions about two types of unobserved heterogeneity. First, we require there be no “persistent unobserved state variables.” That is, we rule out, for example, random and fixed effects in player payoff functions. Second, while the underlying game is permitted to have multiple equilibria, we have to assume a single equilibrium is played in the data. If the first-stage estimates pool across games where multiple equilibria are played, the resulting choice probabilities will be a weighted average of these equilibria, where the weights are given by the equilibrium selection probabilities. Nonlinearity in the second step estimator then leads to inconsistent parameter estimates. Some progress has been made on relaxing these assumptions (see Aguirregabiria and Nevo 2013 for a survey), however, these fixes often only apply to relatively restrictive classes of games and outside these classes custom solutions must be developed. Moreover, implementing these solutions often require increasing the computational complexity and burden of estimation. These considerations suggest that a simple specification test of these assumptions may be an attractive tool for practitioners. The LM test of overdispersion naturally applies here. Our proposed LM test can be viewed as a simple triage tool that allows the researcher to estimate the relatively inexpensive, restrictive model and then use the LM test to decide whether the more complicated model is needed. We demonstrate the performance of our test in detecting failure of the single equilibrium assumption in a Monte Carlo exercise based on the application of Sweeting (2009), who analyzes the timing of radio commercials among rival stations.

2. TEST OF NEGLECTED CORRELATED RANDOM EFFECTS—LM TEST REVISITED

Assume that we observe a random sample \( (Y_i, X_i), i = 1, \ldots, n \). The \( Y \) and \( X \) can be vectors. For example, in the panel data analysis where each individual is observed over \( T \) time periods, we will have \( Y_i = (Y_{i1}, \ldots, Y_{iT}) \) and \( X_i = (X_{i1}, \ldots, X_{iT}) \).

We will assume that the conditional density of \( Y \) given \( X \) is given by the density \( f(y|x, \theta, \gamma) \), where \((\theta, \gamma)\) is the parameter that characterizes the density. Under the null hypothesis, \( \gamma \) is fixed at \( \xi \), but under the alternative hypothesis, the \( \gamma \) may be a random variable potentially correlated with \( X \). Our purpose is to develop an LM test to detect such neglected heterogeneity in \( \gamma \). For simplicity, we will assume that the \( \gamma \) is a scalar. Discussion for the case of multivariate \( \gamma \) can be found in the Appendix B. Following the convention in the literature, for example, Chesher (1984), we will also treat both \( \gamma \) and \( \xi \) as known/ixed, which is the basis of our analysis throughout the rest of this section. Note that we can write \( f(y|x, \theta, \gamma) = f(y|x, \gamma) \) without loss of generality.

We will assume that under the alternative hypothesis, the \( \gamma \) is drawn from a distribution whose density is

\[
s(\gamma|x, \eta) = \frac{1}{\eta} g\left(\frac{\gamma - \xi}{\eta} \bigg| x \right), \tag{1}
\]

where \( \eta > 0 \) and

\[
\int wg(w|x) \, dw \equiv b(x). \tag{2}
\]

Note that the conditional density of \( \gamma \) given \( X \) is allowed to depend on the realization of \( X \).

In the existing literature, including Breusch and Pagan (1980), Chesher (1984), Lee and Chesher (1986), and Hahn, Newey, and Smith (2014), it is (implicitly or explicitly) assumed that (i) \( \gamma \) and \( X \) are independent of each other and (ii) the mean of \( \gamma \) is \( \xi \), so that \( b(x) = 0 \).

Our specification and analysis differ from the existing literature in that both of these restrictions are relaxed. In what follows we consider two cases, (i) when \( b(x) \) is unknown and not necessarily zero and (ii) when \( b(x) = 0 \).

2.1 Case 1: Arbitrary \( b(x) \)

When the conditional density function of \( \gamma \) is \( \frac{1}{\eta} g\left(\frac{\gamma - \xi}{\eta} \bigg| x \right) \) as in (1), the density function of the observation \( Y \) is

\[
h(y|x, \xi, \eta) = \int f(y|x, \gamma) \frac{1}{\eta} g\left(\frac{\gamma - \xi}{\eta} \bigg| x \right) \, d\gamma. \tag{3}
\]
By the change-of-variable $\gamma = \xi + \eta w$, we have

$$h (y | x, \xi, \eta) = \int f (y | x, \xi + \eta w) g (w | x) dw.$$ 

The null hypothesis of interest is that there is no heterogeneity in the parameter $\gamma$. In the model above, this is equivalent to the test of the null $H_0 : \eta = 0$. Our LM test is based on

$$s_\eta (y, x, \xi, \eta) = \frac{\partial h (y | x, \xi, \eta) \partial \eta}{h (y | x, \xi, \eta)}$$

evaluated at $\eta = 0$. We can easily see that at $\eta = 0$, we have

$$s_\eta (y, x, \xi, 0) = \frac{f_\gamma (y | x, \xi)}{f (y | x, \xi)} \int w g (w | x) dw = s_\gamma (y | x, \xi) b (x),$$

where

$$s_\gamma (y | x, \xi) = \frac{f_\gamma (y | x, \xi)}{f (y | x, \xi)}.$$

The analysis leading to (4) implies that the LM test reduces to a test of the moment restriction

$$E [s_\gamma (y | x, \xi) b (x)] = 0.$$ 

If we are to have power against all possible distributions of $\gamma$, we need the above conditional moment restriction to hold for all (square integrable) $b (x)$'s. Obviously, it is equivalent to the test of conditional moment restriction

$$E [s_\gamma (y | x, \xi) | x] = 0.$$ 

We have shown that test of $H_0 : E [s_\gamma (y | x, \xi) | x] = 0$ can be viewed as an LM test of neglected heterogeneity with arbitrary $b (x)$, which does not seem to be known to the profession, at least not explicitly. That being said, we avoid further discussion of how this test should be implemented for two reasons. First, there already exist many articles that show how generic tests of conditional moment restrictions can be implemented. These articles consider a restriction of the form $E [g (z) | x] = 0$ for any given function $g (\cdot)$. A partial list includes Newey (1985), Bierens (1990), and Donald, Imbens, and Newey (2003). Users interested in testing $E [s_\gamma (y | x, \xi) | x] = 0$ can simply use one of the existing tests; all they need to do is to replace $g$ by $s_\gamma$. Given that there is not much we can add to the technical discussion on the generic test of conditional moment restrictions, it is deemed sufficient to point out the existing literature. Second, the test of the conditional moment restriction (6) may be avoided from a practical point of view. We find in Section 2.2.2 that for some nonlinear models, the LM test based on the restriction $b (x) = 0$ does seem to have power even when $b (x) \neq 0$. This seems to indicate that unless the LM test is specifically applied to linear models, a user may settle with the seemingly restrictive test discussed in Section 2.2. Obviously this is reasonable when the user’s sole objective is to detect neglected heterogeneity, and it does not mean that the restriction $E [s_\gamma (y | x, \xi) | x] = 0$ is unimportant in itself.

**Example 1.** Consider the model

$$Y_{it} = y_{it} + X_{it} \theta + \epsilon_{it},$$

where $\epsilon_{it}$ is iid $N (0, \omega^2)$ with $\omega^2 = 1$ known. We have

$$\log f (y | x, \xi) = C - \frac{1}{2} \sum_{i=1}^{T} (y_i - x_i \theta)^2,$$

and the test of interest is of the conditional moment restriction

$$E \left[ \sum_{i=1}^{T} (Y_{it} - x_{it} \theta) \left| X_{i1}, \ldots, X_{iT} \right. \right] = 0.$$ 

**Remark 1.** In Example 1, a simple practical version of the test is a $J$-test for the moment (here we treat $\theta$ and $\xi$ to be known, although in practice, we would plug in a consistent estimators of $\theta$ and $\xi$). Therefore, the critical value in practice would have to reflect the sampling variation of the plugged-in estimators as well as the test statistic itself:

$$E \left[ \begin{bmatrix} X_{it} \vline \cdots \vline \cdots \vline X_{iT} \end{bmatrix} \left( \sum_{i=1}^{T} (Y_{it} - x_{it} \theta) \right) \right] = 0.$$ 

2.2 Case 2: $b (x) = 0$

We now consider the special case, where we impose the restriction that $b (x) = 0$. The restriction leads to the problem that the LM test statistic in (4) becomes identically equal to zero. To tackle this problem, we follow Chesher (1984) and Lee and Chesher (1986), for example, and work instead with $\sqrt{n}$, that is,

$$h (y | x, \xi, \eta) = \int f (y | x, \gamma) \frac{1}{\sqrt{n}} g (\frac{y - \xi}{\sqrt{n}} | x) dy.$$ 

It is straightforward to show that the LM test should be based on the $\sqrt{n}$-standardized sample average of

$$\frac{1}{n} f_\gamma (y | x, \xi) \sigma^2 (x),$$

where

$$\sigma^2 (x) \equiv \int w^2 g (w | x) dw.$$ 

Because the variance of (8) is $\Omega \equiv E [r^2 (X_i) (\sigma^2 (X_i))^2]$, where we define

$$r^2 (X_i) \equiv E \left[ \frac{1}{2} f_\gamma (Y_i | X_i, \xi) \right] | X_i],$$

our test statistic would take the form

$$T_n (\sigma^2) = \left( \frac{1}{n} \sum_{i=1}^{n} r^2 (X_i) (\sigma^2 (X_i))^2 \right)^{-1} \times \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{2} f_\gamma (Y_i | X_i, \xi) \sigma^2 (X_i) \right)^2.$$ 


Under some regularity conditions, it can be shown that (this is a straightforward consequence of Lemma A.1 in Appendix A):

\[ T_n(\sigma^2) \Rightarrow \chi^2(1) \]

as \( n \to \infty \), under the null hypothesis that \( \eta = 0 \).

2.2.1 Max-Min Consideration. The test statistic in (10) is inconvenient because it requires that \( \sigma^2(x) \) be specified. In this section, we provide a justification of using \( \sigma^2(x) = 1 \) from a (local) max-min perspective. For expository simplicity, we assume that \( \xi \) and \( \tau^2(\cdot) \) are known.

Now suppose that nature chooses an alternative density \( g_\ast(w|x) \) such that \( \int w g_\ast(w|x)dw = b_\ast(x) = 0 \) and \( \int w^2 g_\ast(w|x)dw = \sigma^2_\ast(x) > 0 \). Suppose that \( \eta = \frac{1}{\sqrt{n}} \). Then, under some regularity conditions, it can be shown (see Lemma A.1 in Appendix A) that under the sequence of distributions characterized by the density

\[ \int f(y|x, \xi + \sqrt{n}w) g_\ast(w|x) dw, \quad (11) \]

we have

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \frac{1}{2} \frac{f_{Y|Y_i X_i, \xi}}{f(Y_i | X_i, \xi)} \sigma^2_\ast(X_i) \right) \\
\Rightarrow N \left( \frac{E \left[ \left( \frac{1}{2} \frac{f_{Y|Y_i X_i, \xi}}{f(Y_i | X_i, \xi)} \sigma^2_\ast(X_i) \right) \right.} \right), \Omega .
\]

This implies that under the local alternative (11) chosen by nature, the test statistic \( T_n \) in (10) converges to a noncentral \( \chi^2 \) distribution with the noncentrality parameter equal to

\[
\phi \left( \sigma^2(\cdot), \sigma^2_\ast(\cdot) \right) = \frac{E \left\{ \tau^2(X_i) \sigma^2_\ast(X_i) \right\}^2}{E \left\{ \tau^2(X_i) \sigma^2(X_i) \right\}^2}.
\]

Suppose that \( c \) is the asymptotic critical value of the test statistic \( T_n(\sigma^2) \). Suppose that \( P_{n, \sigma} \) denotes the probability measure with density equal to (11) with \( \int w^2 g_\ast(w|x)dw = \sigma^2_\ast(x) \) chosen by nature. Then, the asymptotic local power of the test statistic \( T_n(\sigma^2) \), \( \lim_{n \to \infty} P_{n, \sigma}(T_n(\sigma^2) \geq c) \), is determined by the noncentrality parameter \( \phi(\sigma^2(\cdot), \sigma^2_\ast(\cdot)) \) that depends only on \( \sigma^2(\cdot) \) and \( \sigma^2_\ast(\cdot) \).

Our max-min justification is given in the following result:

**Proposition 1.** Consider the problem

\[ \max_{\sigma^2(\cdot)} \min_{\sigma^2_\ast(\cdot)} \phi \left( \sigma^2(\cdot), \sigma^2_\ast(\cdot) \right), \]

where \( \sigma^2(\cdot) \) and \( \sigma^2_\ast(\cdot) \) are normalized such that \( E[\tau^2(X_i)\sigma^2_\ast(X_i)] = E[\tau^2(X_i)\sigma^2(X_i)] = 1 \). The solution is such that \( \sigma^2_\ast(\cdot) \) is a constant function.

**Proof.** In Appendix A. \( \square \)

The decision problem we consider is that for given \( \sigma^2(\cdot) \), nature chooses the alternative \( \sigma^2_\ast(\cdot) \) to minimize the local power, that is, \( \phi(\sigma^2(\cdot), \sigma^2_\ast(\cdot)) \), and then, the econometrician chooses \( \sigma^2_\ast(\cdot) \) by maximizing the local power, that is, \( \phi(\sigma^2_\ast(\cdot), \sigma^2(\cdot)) \). The result above establishes that the econometrician’s choice is such that \( \sigma^2_\ast(\cdot) \) is constant. In other words, a constant choice for \( \sigma^2(\cdot) \) is optimal from a max-min decision perspective.

Our discussion above suggests that it would be reasonable to take \( \sigma^2(X_i) = 1 \) and can be justified by the max-min principle. Assuming that \( \tau^2(X_i) > 0 \) almost everywhere, we can see that the \( \phi(1, \sigma^2_\ast(X_i)) > 0 \) for any \( \sigma^2_\ast(X_i) \) as long as \( \sigma^2_\ast(X_i) > 0 \) with positive probability. Given that the probability \( \sigma^2_\ast(X_i) > 0 \) is positive for any reasonable \( \sigma^2(X_i) \), we can conclude that the test statistic \( T_n(\sigma^2) \) with \( \sigma^2_\ast(X_i) = 1 \) has a local power against any reasonable alternatives. Proposition 1 has an intuitive interpretation that the random effects specification has the max-min property from a testing perspective. The local power only depends on the conditional second moment and therefore, the random effects specification where \( \gamma \) is independent of \( x \) can be argued to be the optimal specification.

2.2.2 Additional Consideration. We ask whether the test of overdispersion has any power against the local alternative where the mean independence assumption may be violated. Test of the conditional moment restriction (6) is not expected to have any power if the alternative satisfies the mean independence assumption \( b(x) = 0 \). Although the LM test of overdispersion is developed under the assumption that \( b(x) = 0 \), if it has a power even when \( b(x) \neq 0 \), we may argue that it is more robust than the test of conditional restriction. We provide two examples, and discuss the power of the LM test of overdispersion.

Given the max-min consideration, we will only consider the case where \( \sigma^2(\cdot) = 1 \), and examine the power of the test statistic (10) under the local alternative where the mean independence may be violated:

**Proposition 2.** Under the sequence of distributions characterized by the density (3) with \( \eta = \frac{1}{\sqrt{n}} \), the test statistic (10) is asymptotically noncentral \( \chi^2 \) with the noncentrality parameter equal to \( E[\nu(X_i)b(X_i)]^2 E[\tau^2(X_i)] \), where

\[
\nu(X_i) = E \left\{ \left( \frac{1}{2} \frac{f_{Y|Y_i X_i, \xi}}{f(Y_i | X_i, \xi)} \right) \left( \frac{f_{Y|Y_i X_i, \xi}}{f(Y_i | X_i, \xi)} - 1 \right) \right\} X_i.
\]

**Proof.** Using the same argument as in Appendix A with \( \kappa = n^{-1/2} \), we can show that \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{2} \sum_{k=1}^{2} f_{Y|Y_i X_i, \xi} \) is asymptotically normal with mean equal to \( E[\nu(X_i)b(X_i)] \) and variance equal to \( E[\tau^2(X_i)^2] \), from which the conclusion follows. \( \square \)

Below we present two examples, one of which is such that the local power is zero because \( \nu(X_i) = 0 \), and the other one is such that it is not necessarily the case.

**Example 2 (Example 1, continued).** The textbook linear model in Example 1 is such that we have

\[
\log f(Y_i | X_i, \xi) = C - \frac{1}{2} \sum_{t=1}^{T} (Y_{it} - X_{it}^\prime \theta - \xi)^2,
\]

\[
\frac{f_{Y|Y_i X_i, \xi}}{f(Y_i | X_i, \xi)} = \left( \sum_{t=1}^{T} (Y_{it} - X_{it}^\prime \theta - \xi)^2 \right)^{-T} T
\]

\[
= \left( \sum_{t=1}^{T} e_{it}^2 \right)^{-T} T
\]

\[
\frac{f_{Y|Y_i X_i, \xi}}{f(Y_i | X_i, \xi)} = \sum_{t=1}^{T} (Y_{it} - X_{it}^\prime \theta - \xi) = \sum_{t=1}^{T} e_{it}
\]

\[
= \sum_{t=1}^{T} (Y_{it} - X_{it}^\prime \theta - \xi) = \sum_{t=1}^{T} e_{it}
\]
and using the symmetry of the normal distribution, we obtain \( v(X_i) = 0 \).

**Example 3 (Panel Logit).** Consider the panel logit model where

\[
\begin{align*}
L(Y_i | X_i, \xi) &= \prod_{t=1}^{T} \left( \frac{\exp(X_{it}'\theta + \xi)}{1 + \exp(X_{it}'\theta + \xi)} \right)^{Y_{it}} \\
&\quad \times \left( \frac{1}{1 + \exp(X_{it}'\theta + \xi)} \right)^{1-Y_{it}}.
\end{align*}
\]

We can see that

\[
\begin{align*}
\log L(Y_i | X_i, \xi) &= \sum_{t=1}^{T} Y_{it} (X_{it}'\theta + \xi) \\
&\quad - \sum_{t=1}^{T} \log \left( 1 + \exp(X_{it}'\theta + \xi) \right),
\end{align*}
\]

\[
\begin{align*}
\frac{f_{Y}}{f(Y_i | X_i, \xi)} &= \sum_{t=1}^{T} \left( Y_{it} - \frac{\exp(X_{it}'\theta + \xi)}{1 + \exp(X_{it}'\theta + \xi)} \right),
\end{align*}
\]

\[
\frac{f_{YY}}{f(Y_i | X_i, \xi)} = \left( \sum_{t=1}^{T} \frac{\exp(X_{it}'\theta + \xi)}{1 + \exp(X_{it}'\theta + \xi)} \right)^2
\]

\[
\begin{align*}
&- \sum_{t=1}^{T} \frac{\exp(X_{it}'\theta + \xi)}{1 + \exp(X_{it}'\theta + \xi)},
\end{align*}
\]

and

\[
\begin{align*}
\nu(X_i) &= \frac{1}{2} \sum_{t=1}^{T} \frac{\exp(X_{it}'\theta + \xi)(1 - \exp(X_{it}'\theta + \xi))}{(1 + \exp(X_{it}'\theta + \xi))^3}.
\end{align*}
\]

### 3. MONTE CARLO

A potentially important application of our test arises in the context of estimating static binary choice models with social and strategic interactions. These models have been employed to analyze a wide variety of applications including firm entry (Seim 2006), the timing of radio commercials (Sweeting 2009), labor force participation (Bjorn and Vuong 1984), teen sex (Card and Giuliano 2013), and many more. These models can be conveniently estimated by a two-step procedure. In the first step, the economist estimates the so-called *conditional choice probabilities* (CCP). These CCPs are then used as an ingredient in the second step to estimate the structural parameters. (The basic two-step structure of this model was pioneered by Hotz and Miller (1993) in the context of dynamic discrete choice models and has subsequently been extended to static discrete games (Aguirregabiria and Mira 2003; Bajari et al. 2010) and dynamic games (Aguirregabiria and Mira 2007; Bajari, Benkard, and Levin 2007; Pakes, Ostrovsky, and Berry 2007; Pesendorfer and Schmidt-Dengler 2008).) The first-stage CCP estimates require pooling of observations across games and agents, which leads to two sources of potential neglected heterogeneity that may invalidate the approach. First, unobserved individual or game/group level payoff heterogeneity and second, multiple equilibria being played across games. Both of these make the CCP inconsistent estimates of equilibrium choice probabilities and, in turn, lead to inconsistency of the second-stage estimates. We will illustrate how the LM test statistic can be used to detect such problems, exploiting the intuition by, for example, Bajari et al. (2011) that multiple equilibria in games can be viewed as a particular from of unobserved heterogeneity.

We start with a brief description of the model. Assume that the econometrician has access to data \(\{D_{ig}, X_{ig}\}_{i=1}^{n} \) on binary actions, \(D_i\), and covariates, \(X_i\) for \(K_g\) agents interacting in each of \(n\) games. The objective of estimation is to estimate payoffs \(u(d_{ig}, d_{-ig}, x_{ig}; \theta)\) that rationalize observed choice behavior, where \(D_{-ig}\) is the vector of choices of all agents except \(i\). We assume utilities depend on the sample average of choices of other players and covariates in the following linear, separable way:

\[
\begin{align*}
u(d_{ig}, d_{-ig}, x_{ig}; \theta) &= d_{ig} \left( B x_{ig} + J_{1} g, -1 \sum_{j \neq i} d_{jg} \right) \\
&\quad + (1 - d_{ig}) \left( C x_{ig} + J_{2} g, -1 \sum_{j \neq i} d_{jg} \right).
\end{align*}
\]

Following Bajari et al. (2010), we assume the data are generated by a Bayes–Nash equilibrium of a game of incomplete information where players choose actions, \(d_{ig} \in [0, 1]\) to maximize

\[
E_{D_{ig}}[u(d_{ig}, D_{-ig}, x_{ig}; \theta) + \varepsilon_{d_{ig}} | X_g = x_g]
\]

\[
= d_{ig} \left( B x_{ig} + J_{1} g, -1 \sum_{j \neq i} D_{jg} \right| X_g = x_g) + \varepsilon_{1ig}
\]

\[
+(1 - d_{ig}) \left( C x_{ig} + J_{2} g, -1 \sum_{j \neq i} D_{jg} \right| X_g = x_g)
\]

\[
+ \varepsilon_{0ig}
\]

where \(\varepsilon_{0ig}(d_{ig})\) is an iid, across agents and games, private information payoff shock drawn from a commonly known distribution and we use the convention \(X_g = (X_{ig}, \ldots, X_{K_gg})\), and the probability \(f(d_{ig} = 1|x_{ig}; \theta)\) that player \(i\) chooses \(d_{ig} = 1\) is

\[
Pr(\varepsilon_{0ig} - \varepsilon_{1ig} \leq E_{D_{ig}}[u(1, D_{-ig}, x_{ig}; \theta)] - u(0, D_{-ig}, x_{ig}; \theta) | X_g = x_g) \leq 0.16.
\]

The structure above leads to a natural two-step estimator for \(\theta\) as discussed in the beginning of this section, using (ideally nonparametric) estimates of the CCP (in the common usage the CCP for this example would be the individual conditional choice probability \(E[D_{ig}|X_{ig}]\). Here, the assumption that payoffs only depend on average rival actions and these enter only linearity in the payoff function means the scalar, conditional average choice function, (13) is sufficient for the vector of individual choice probabilities:

\[
E \left[ \frac{1}{K_g - 1} \sum_{j \neq i} D_{jg} \right| X_g = x_g = \xi(x_g).
\]

(13)
When there are multiple equilibria, \( E(\frac{1}{K} \sum_{j \neq i} D_{jg} | X_g = x_g) \) can be viewed as a random variable \( \gamma(x_g) \); the moment Equation (13) defines a pseudo-parameter \( \xi(x) \), which is an average of the \( \gamma(x) \) weighted by the probability of equilibrium selection. It follows that the LM test statistic of interest in this application is simply

\[
\frac{1}{\sqrt{n}} \sum_{g=1}^{n} \sum_{i=1}^{K} f_{gY}(D_{ig} | X_{ig}, \theta, \xi(X_g)) \cdot f(D_{ig} | X_{ig}, \theta, \xi(X_g))
\]

Obviously, one needs to adjust for the error of estimating \( \theta \) as well as \( \xi(X_g) \). The \( \xi(X_g) \) is often estimated nonparametrically, and as such, analysis along the lines of Newey (1994) is necessary to characterize the asymptotic variance. From a practical perspective, it can be accommodated either by the bootstrap result as in Chen, Linton, and van Keilegom (2003), or by the numerical equivalence result as in Ackerberg, Chen, and Hahn (2012). (In a supplementary appendix available upon request, we present a (straightforward) characterization of how the procedure in Ackerberg, Chen, and Hahn (2012) can be used in our context.)

In principle, one may want to consider Wald or LR tests as well, using the recent progress on identification and two-step estimation of models with multiple equilibria and unobserved payoff heterogeneity, in particular, by Kasahara and Shimotsu (2009). However, in contrast to our LM test, both the Wald and LR tests require the econometrician to be able to specify the number of equilibria/types. Both full solution approaches, where the economist actually solves for equilibrium for every parameter guess in the estimation search algorithm, and two-step approaches admitting multiple equilibria/payoff heterogeneity entail a substantial increase in the computational burden of estimation. In this sense, our LM test can be looked at as a triage tool that allows the researcher to judge whether incurring such costs is necessary. Of course, any such pretesting strategy should account for the inherent potential distortions. Formal treatment of this problem is outside the scope of this article and any actual implementation will depend on the context of the application so we do not address these issues further here.

Our Monte Carlo exercise is inspired by the application of Sweeting (2009) discussed by Sweeting (2009) discussed how “overdispersion” in the distribution of the CCP can be exploited for identification in the context of the two-step estimator, although he does not employ this approach and argues informally that the data support the single equilibrium assumption. Sweeting did develop a formal modified log-likelihood test for the nested fixed point estimator, but such a test cannot be developed for the two-step estimator because the CCP are not identified under the alternative of multiple equilibria, who estimated a strategic model of the timing of radio commercials by radio stations. There may be incentives for radio stations to coordinate the timing of their commercials so listeners do not have incentive to switch stations when commercials come on because, if they do, they will just tune to another commercial. We consider the simplified version of the model discussed earlier, and assume there are \( K \) stations competing in each of \( n \) games. For ease of exposition, we assume the players are local in the sense that each player only competes in 1 game. Stations compete by choosing a binary action \( d_{ig} \in \{0, 1\} \), which can be interpreted as a choice to play commercials at :50 or :55 min past the hour. The payoffs associated with each choice are special cases of the payoffs presented above and are given by

\[
E[u(d_{ig} = 1, D_{-ig}, x_g, \theta)] = X_{ig} + \epsilon_{1ig}
\]

\[
= BX_{ig} + J\left(\sum_{j \neq i} \frac{d_{jg}}{K-1} | X_g, \theta, \xi(X_g)\right) + \epsilon_{1ig}
\]

\[
E[u(d_{ig} = 0, D_{-ig}, x_g, \theta)] = X_{ig} + \epsilon_{0ig}
\]

\[
= CX_{ig} + J\left(\sum_{j \neq i} \frac{1 - d_{jg}}{K-1} | X_g, \theta, \xi(X_g)\right) + \epsilon_{0ig}.
\]

We assume \( X_{ig} \in \{-1, 0, 1\} \) and are drawn iid uniform across players and games. Finally, we assume the \( \epsilon \)'s are drawn iid from a Type 1 extreme value distribution.

The probability of observing \( d_{ig} = 1 \) is given by

\[
Pr(d_{ig} = 1 | X_g, \theta) = Pr(\tilde{d}(d = 1, X; \theta) > 0)
\]

\[
= \exp(\beta X_{ig} + J(2E\left(\sum_{j \neq i} \frac{d_{jg}}{K-1} | X_g, \theta\right) - 1))
\]

\[
1 + \exp(\beta X_{ig} + J(2E\left(\sum_{j \neq i} \frac{d_{jg}}{K-1} | X_g, \theta\right) - 1)).
\]

where \( \beta = (B - C) \) and where we let \( \tilde{d}(d = d, X; \theta) = E[u(d = d, D_{-ig}, x_g, \theta)] | X_g = x_g \]. In the second stage, we formulate a (pseudo) maximum likelihood estimator based on the logit conditional choice probabilities:

\[
f(d | X, \theta, \hat{P}(X)) = \left(\frac{\exp(\beta X_{ig} + J(\hat{\beta}(X_{ig}' - 1)) - 1)}{1 + \exp(\beta X_{ig} + J(\hat{\beta}(X_{ig}' - 1))}ight)^{d_{ig}}
\]

\[
\times \left(\frac{1}{1 + \exp(\beta X_{ig} + J(\hat{\beta}(X_{ig}' - 1))}ight)^{1 - d_{ig}}
\]

where we have used exchangeability of the \( X \)'s to rearrange \( X_g \) as \( X_{ig}' = (X_{ig}, X_{-ig}) \) and drop the \( i \) index on the CCP and \( \hat{P} \) is the first-stage CCP estimate:

\[
\hat{P}(X_{ig}') = E\left[\sum_{j \neq i} \frac{d_{jg}}{K-1} | X_g, \theta\right]
\]

\[
= \frac{1}{\pi K} \sum_{g=1}^{n} \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{1}{K-1} \sum_{k=1}^{K} \sum_{j=1}^{K} \sum_{i=1}^{K} \sum_{g=1}^{n} |X_{ig}' = x_{ig}'|.
\]

Letting \( \hat{\theta} = (\hat{\beta}, \hat{J}) \) denote the maximum likelihood estimates, our test statistic is

\[
\hat{u} = \frac{1}{\sqrt{n}} \sum_{g=1}^{n} \sum_{i=1}^{K} f_{gY}(d_{ig} | X_{ig}, \hat{\theta}, \hat{P}_{ig}) \frac{1}{\sqrt{n}} \sum_{g=1}^{n} \sum_{i=1}^{K} |d_{ig} = 1|
\]

\[
\times \left(\frac{1 + \exp(\hat{\beta} X_{ig} + \hat{J}(\hat{\beta}(X_{ig}' - 1)) - 1)}{1 + \exp(\hat{\beta} X_{ig} + \hat{J}(\hat{\beta}(X_{ig}' - 1)) - 1)}
\]

\[
-4J^2 \times \left(\frac{1 + \exp(\hat{\beta} X_{ig} + \hat{J}(\hat{\beta}(X_{ig}' - 1)) - 1)}{1 + \exp(\hat{\beta} X_{ig} + \hat{J}(\hat{\beta}(X_{ig}' - 1)) - 1)}
\]

\[
\times \left(\frac{1 + \exp(\hat{\beta} X_{ig} + \hat{J}(\hat{\beta}(X_{ig}' - 1)) - 1)}{1 + \exp(\hat{\beta} X_{ig} + \hat{J}(\hat{\beta}(X_{ig}' - 1)) - 1)}
\]

\[
\right)^3.
\]
NOTES: Fraction of 2000 simulated datasets where the null (a single equilibrium is played for each covariate configuration) is rejected at the indicated significance level. Test statistic distribution estimated using 250 bootstrap repetitions. n is the number of observed games, where each game has five players. As described in the text, each game has five players, payoff shifters, \( X_g \), are drawn uniform iid across players and games. Equilibrium selection is determined by iteratively updating best responses starting from a randomly selected deterministic strategy profile conditional on: (1) \( X \) under the null and (2) \( X_g \) under the alternative.

Table 1. Parameter estimates

| \( J \) | \( n \) | \( \hat{\beta}_0 \) (const.; \( \beta_0 = 0 \)) | \( \hat{\beta}_1 (\beta_1 = 1) \) | \( J \) |
|-------|-------|-----------------|-----------------|-------|
| 2.75  | 100   | 0.002 (0.09)    | 1.036 (0.23)    | 2.818 (0.13) |
|       | 200   | 0.002 (0.06)    | 1.012 (0.16)    | 2.784 (0.09) |
|       | 500   | 0.001 (0.04)    | 1.009 (0.10)    | 2.762 (0.06) |
|       | 100   | 0.003 (0.09)    | 1.048 (0.26)    | 3.088 (0.18) |
| 3     | 200   | 0.003 (0.06)    | 1.026 (0.18)    | 3.049 (0.12) |
|       | 500   | −0.003 (0.04)   | 1.016 (0.10)    | 3.024 (0.07) |
|       | 100   | 0.007 (0.10)    | 1.074 (0.29)    | 3.357 (0.22) |
| 3.25  | 200   | 0.002 (0.07)    | 1.042 (0.20)    | 3.311 (0.15) |
|       | 500   | 0.000 (0.040977)| 1.008 (0.12)    | 3.268 (0.08) |

NOTES: Average maximum likelihood coefficient estimates for 2000 simulated datasets, standard deviations of the simulated estimates are in parentheses. \( n \) is the number of observed games, where each game has five players. As described in the text, each game has five players, payoff shifters, \( X_g \), are drawn uniform iid across players and games. Equilibrium selection is determined by iteratively updating best responses starting from a randomly selected deterministic strategy profile conditional on: (1) \( X \) under the null and (2) \( X_g \) under the alternative.

\[
1 + [d_{ig} = 0] \left( 1 + \exp(\hat{\beta}_X_i g + \hat{J}(2\hat{P}(X_i g) - 1)) \right) \times \left( 1 - \exp(\hat{\beta}_X_i g + \hat{J}(2\hat{P}(X_i g) - 1)) \right) \times \exp(\hat{\beta}_X_i g + \hat{J}(2\hat{P}(X_i g) - 1)).
\]

For our Monte Carlo experiments, we set \( K = 5 \). We first solve for all the stable equilibria of our game conditional on \( X_g \), then we generate data assuming the \( X_g \)'s are drawn from a discrete uniform distribution in the population. To generate choice data under the single equilibrium assumption, we randomly select an equilibrium from the calculated set of equilibria for each realization of \( X_g \) in the data and then generate choices by drawing from the distribution induced by this equilibrium. In both the case where the single equilibrium assumption holds and where it does not, we use the same equilibrium selection mechanism. To define the mechanism, we start by using iterative updating of strategies to find an equilibrium associated with an initial starting strategy profile that has all players choosing an action with probability one. There are 32 (= 2^5) such profiles and all converge to one of the, at most, two stable equilibria of the game. For each \( X_g \), in the case of the single equilibrium assumption holding, and for each \( g \), in the case of the single equilibrium assumption failing, the selection mechanism chooses uniformly from the 32 equilibrium realizations of one of the, at most, two equilibria found using our iterative updating algorithm. Intuitively, this causes the mechanism to, for example, tend to select the “high” equilibrium more often when covariate values tend to be higher. Defining the selection mechanism in this way is a natural way to generate dependence of the equilibrium selection on the covariates.

Tables 1 and 2 show the results of our Monte Carlo experiments. Table 1 displays the mean and standard deviation of

Table 2. Rejection frequency

| \( J \) | \( n \) | Reject (\( p = 0.1 \)) | Reject (\( p = 0.05 \)) | Reject (\( p = 0.01 \)) | Reject (\( p = 0.1 \)) | Reject (\( p = 0.05 \)) | Reject (\( p = 0.01 \)) |
|-------|-------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 2.75  | 100   | 0.061           | 0.027           | 0.005           | 0.28            | 0.179           | 0.056           |
|       | 200   | 0.070           | 0.036           | 0.005           | 0.362           | 0.257           | 0.107           |
|       | 500   | 0.080           | 0.029           | 0.006           | 0.635           | 0.518           | 0.279           |
|       | 100   | 0.034           | 0.012           | 0.001           | 0.442           | 0.314           | 0.125           |
| 3     | 200   | 0.051           | 0.022           | 0.002           | 0.715           | 0.566           | 0.306           |
|       | 500   | 0.079           | 0.034           | 0.01            | 0.955           | 0.921           | 0.788           |
|       | 100   | 0.029           | 0.012           | 0.001           | 0.717           | 0.538           | 0.246           |
| 3.25  | 200   | 0.046           | 0.018           | 0.001           | 0.956           | 0.867           | 0.543           |
|       | 500   | 0.076           | 0.031           | 0.003           | 0.999           | 0.999           | 0.987           |

NOTES: Fraction of 2000 simulated datasets where the null (a single equilibrium is played for each covariate configuration \( X \)) is rejected at the indicated significance level. Test statistic distribution estimated using 250 bootstrap repetitions. \( n \) is the number of observed games, where each game has five players. As described in the text, each game has five players, payoff shifters, \( X_g \), are drawn uniform iid across players and games. Equilibrium selection is determined by iteratively updating best responses starting from a randomly selected deterministic strategy profile conditional on: (1) \( X \) under the null and (2) \( X_g \) under the alternative.
parameter estimates from simulations assuming different sample sizes and true underlying interaction effects. Table 2 explores the power and size attributes of the test, displaying rejection probabilities from simulations across the different sample sizes and true interaction effects. For each simulation, we calculate

\[ \hat{\alpha} = \frac{\hat{\sigma}}{\text{var}(\hat{\sigma})}, \]

where we use 250 bootstrap to estimate \text{var}(\hat{\sigma}), and compare it to the standard normal critical values. For both tables, the statistics are based on 2000 simulations.

The last three columns of Table 1 illustrate the impact of the failure of the single equilibrium assumption. Both the impact of fundamentals \((\hat{\beta}_t)\) and strategic effects \((J)\) are attenuated. The reason for this is simple. Due to multiplicity, actual choice probabilities are either all high or all low depending on which equilibrium is selected. The CCP estimate lies in between as a selection-probability weighted average of these equilibria. In other words, our CCP estimates suggest that equilibrium choices are more independent than they really are and thus more consistent with a game with relatively low strategic interaction effects.

Table 2 shows our multiplicity test performs quite well in this example. When strategic effects are weaker, variation in outcomes becomes decreasingly dominated by multiplicity and our test has less power in detecting a violation of the single equilibrium assumption. In our example, this happens both because the direct effect of \(X_t\) on a player’s payoffs is larger relative to the strategic effects and also indirectly because the equilibrium selection mechanism is relatively more dependent on \(X_g\).

**APPENDIX A: TECHNICAL DETAILS FOR SECTION 2**

A.1 Regularity Conditions and Preliminary Result

We first provide regularity conditions.

**Assumption A.1.** Assume that \((Y_t, X_t)\) are i.i.d.

We denote \(p(x)\) as the density function of \(X_t\) and \(X\) as the support of \(X_t\).

**Assumption A.2.** We assume that the conditional density of \(w\) on \(x, g_t (w|z)\), satisfies the following conditions. (i) \(\int w g_t (w|z) dw = b_t (x) = 0\), (ii) \(\int w^2 g_t (w|z) dw = \sigma^2 (x) > 0\), and (iii) \(\int |w|^k g_t (w|z) dw \leq G_k (x)\) for all \(x \in X\).

**Assumption A.3.** We assume that (i) the conditional density \(f (y|x, \gamma)\) is twice differentiable with respect to the argument \(\gamma\), (ii) there exists a function \(F (y, x)\) such that \(\sup_y |f (y|x, \gamma)|, \sup_y |f (y|x, \gamma)|, \sup_{y, \gamma} |J (y|x, \gamma)| \leq F (y, x)\), and \(\left| f_{x \gamma} (y|x, \gamma_1) - f_{x \gamma} (y|x, \gamma_1) \right| \leq F (y, x)|\gamma_1 - \gamma_1|\) with \(F (y, x) dy < \infty\) for all \(x \in X\) and satisfying the following: (ii)-a) \(\int \int \sigma^2 (x) \left| f_{x \gamma} (y|x, \gamma) \right|^2 f (y|x, X_t) G_t (x) p (x) dy dx < \infty\), (ii)-b) \(\int \int \sigma^2 (x) \left| f_{x \gamma} (y|x, \gamma) \right|^2 f (y|x, X_t) G_t (x) p (x) dy dx < \infty\), and (ii)-c) \(\int \int \sigma^2 (x) \left| f_{x \gamma} (y|x, \gamma) \right|^2 f (y|x, X_t) G_t (x) p (x) dy dx < \infty\) for all \(x \in X\), where \(G_t (x)\) is the bound of Assumption A.2(ii).

A.2 Limit \(T_n (\hat{\alpha}^2)\) in (10) Under Local Alternatives

We prove that the test statistic based on (8), that is,

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( \frac{1}{2} f_{y} (Y_t | X_t, \xi) \sigma^2 (X_t) \right) \]

is asymptotically normal with mean equal to

\[ E \left[ \sigma^2 (X_t) \sigma^2 (X_t) \right] \]

and variance equal to

\[ \Omega = E \left[ \sigma^2 (X_t) \right] \]

under the local alternative \(k (y|x, \xi; \kappa) = \int f (y|x, \xi + \kappa w) g_t (w|x) dw\).

**Lemma A.1.** Assume Assumptions A.1–A.2. Then, under the sequence of distributions characterized by the density

\[ \int f (y|x, \xi + \sqrt{\eta} w) g_t (w|x) dw \]

with \(\eta = \frac{1}{\sqrt{\sigma}}\), the test statistic \(T_n (\sigma^2)\) in (10) converges to a noncentral \(\chi^2\) distribution with the noncentrality parameter equal to

\[ \phi \left( \sigma^2 (\cdot), \sigma_k^2 (\cdot) \right) = \left( E \left[ \sigma^2 (X_t) \sigma^2 (X_t) \right] \right)^2. \]

We denote \(P_\kappa, E_\kappa,\) and \(V_\kappa\) as the probability measure, the expectation operator, and the variance operator, respectively, defined by the density \(k (y|x, \xi; \kappa) p (x)\). Define \(P_\kappa = P_0, E_\kappa = E_0,\) and \(V_\kappa = V_0\).

We first derive the limits

\[ \lim_{\kappa \to 0} E_\kappa \left[ \frac{1}{2} f_{y} (Y_t | X_t, \xi) \right] \]

and

\[ \lim_{\kappa \to 0} V_\kappa \left[ \frac{1}{2} f_{y} (Y_t | X_t, \xi) \right]. \]

Part (i): We apply the second-order Taylor approximation to \(k (y|x, \xi; \kappa) = f (y|x, \xi)\) as a function of \(\kappa\) around \(\kappa = 0\). Then,

\[ \frac{\partial k (y|x, \xi; 0)}{\partial \kappa} + \frac{1}{2} \left( \frac{\partial^2 k (y|x, \xi; 0)}{\partial \kappa^2} \right) \kappa^2 + \left( \frac{\partial^2 k (y|x, \xi; \kappa)}{\partial \kappa^2} \right) \kappa^2 \]

where \(\xi\) lies between 0 and \(\kappa\). Under Assumptions A.3 and A.2, we have

\[ \frac{\partial k (y|x, \xi; 0)}{\partial \kappa} = \int \left( \frac{\partial f (y|x, \xi + \kappa w)}{\partial \kappa} \right)_{\kappa=0} w g_t (w|x) dw = b_t (x) = 0, \]

\[ \frac{\partial^2 k (y|x, \xi; 0)}{\partial \kappa^2} = \int \left( \frac{\partial^2 f (y|x, \xi + \kappa w)}{\partial \kappa^2} \right)_{\kappa=0} w g_t (w|x) dw = \frac{1}{2} f_{y} (y|x, \xi) \sigma_k^2 (x), \]
Then, we have

\[
\begin{align*}
\left| \frac{\partial^2 k(y|x, \xi; \tilde{k})}{\partial^2 \kappa} - \frac{\partial^2 f(y|x, \xi; 0)}{\partial^2 \kappa} \right|_k &= \left| \int \left\{ \frac{\partial^2 f(y|x, \xi + \kappa w)}{\partial^2 \kappa} \right\} \times w^2 g_s(w|x)dw \right|_k = \left| \int \left\{ \frac{\partial^2 f(y|x, \xi + \kappa w)}{\partial^2 \kappa} \right\} \times w^2 g_s(w|x)dw \right|_k \\
&\leq \int \left| \frac{\partial^2 f(y|x, \xi + \kappa w)}{\partial^2 \kappa} \right| \times w^2 g_s(w|x)dw \\
&\leq \int |f_{yy}(y|x, \xi) - f_{yy}(y|x, \xi + \tilde{k} w)|w^2 g_s(w|x)dw \\
&\leq \kappa F(y, x) \int |w|^2 g_s(w|x)dw \\
&\leq \kappa F(y, x) G_s(x). \quad (A.3)
\end{align*}
\]

Choose

\[ \kappa = n^{-1/4}, \quad (A.4) \]

Then, we have

\[
\begin{align*}
E_s \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{f(Y_i|X_i, \xi)} \sigma^2(X_i) \right] &= \sqrt{n} \int \left[ \frac{1}{f(Y|X, \xi)} \sigma^2(x) k(y|x, \xi; \kappa) p(x) dy dx \right] \\
&= \sqrt{n} \int \left[ \frac{1}{f(Y|X, \xi)} \sigma^2(x) \left( \frac{1}{f(Y|X, \xi)} \sigma^2(x) \right) \right] \times p(x) dy dx \\
&= \int \int \left[ \frac{1}{f(Y|X, \xi)} \sigma^2(x) \right] \left[ \frac{1}{f(Y|X, \xi)} \sigma^2(x) \right] \times p(x) dy dx \\
&+ \int \int \left[ \frac{1}{f(Y|X, \xi)} \sigma^2(x) \right] \left| \frac{\partial^2 k(y|x, \xi; \tilde{k})}{\partial^2 \kappa} - \frac{\partial^2 f(y|x, \xi; 0)}{\partial^2 \kappa} \right| p(x) dy dx \quad (A.5)
\end{align*}
\]

where the first equality holds by Assumption A.1 and the second equality holds since \( f_{yy}(y|x, \xi) dy = 0 \) under Assumption A.3, and the third equality holds by the second-order Taylor approximation under Assumption A.3. Notice that by (A.3) and (A.4), we have

\[
\begin{align*}
E_s \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{f(Y_i|X_i, \xi)} \sigma^2(X_i) \right] &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{f(Y_i|X_i, \xi)} \sigma^2(X_i) \\
- \int \int \left[ \frac{1}{f(Y|X, \xi)} \sigma^2(x) \right] \left[ \frac{1}{f(Y|X, \xi)} \sigma^2(x) \right] f(y|x, \xi; \kappa) p(x) dy dx \\
&\leq \frac{1}{2} \int \sigma^2(x) \left| \frac{1}{f(Y|X, \xi)} \right|^2 p(x) dy dx \\
&\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{f(Y_i|X_i, \xi)} \sigma^2(X_i) \\
&\leq n^{-1/4} M(\xi)
\end{align*}
\]

and consider a triangular sequence \( U_i : 1 \leq i \leq n \) and its underlying probability measure sequence \( P_{n^{-1/4}} \). Since \( \{U_i\}_{i=1, \ldots, n} \sim \text{iid}(0, 1) \) under \( P_{n^{-1/4}} \), by the Lindberg–Feller central limit theorem (CLT) for the triangular array, we have

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \Rightarrow N(0, 1) \text{ under } P_{n^{-1/4}}. \quad (A.6)
\]
By Parts (i) and (ii) and (A.6), under $P_{n-1/4}$ we have
\[
\begin{align*}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{2} f_{Y_i}(Y_i, x_i, \xi) \sigma^2(X_i) \\
= & \left( V_{n-1/4} \frac{1}{2} f_{Y_i}(Y_i, x_i, \xi) \sigma^2(X_i) \right)^{1/2} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \right) \\
+ & E_{n-1/4} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{1}{2} f_{Y_i}(Y_i, x_i, \xi) \sigma^2(X_i) \right] \\
\Rightarrow & \Omega^{1/2} N(0, 1) + \left( \frac{1}{2} f_{Y_i}(Y_i, x_i, \xi) \sigma^2(X_i) \right) \text{under } P_{n-1/4} \\
= & N \left( E \left[ \frac{1}{2} f_{Y_i}(Y_i, x_i, \xi) \sigma^2(X_i) \right] \right) \Omega.
\end{align*}
\]

as required for Lemma A.1.

\square

### A.3 Proof of Proposition 1

The nature moves first. To minimize the power, it tries to minimize the noncentrality parameter $\phi(\sigma^2(\cdot), \sigma^2(\cdot))$ by choosing $\sigma^2(x)$ given positivity of $\sigma^2(\cdot) > 0$ and the normalization restriction and $\sigma^2(\cdot)$. This is equivalent to minimizing $E[\tau^2(X_i) \sigma^2(X_i)]$ with respect to $\sigma^2(x)$ given positivity of $\sigma^2(\cdot) > 0$ and the normalization restriction. Let $\mathcal{X}$ denote the support of $X_i$. Notice that for any $\sigma^2(x)$ with $E[\tau^2(X_i) \sigma^2(X_i)] = 1$, it follows that
\[
E[\tau^2(X_i) \sigma^2(X_i)] = \min_{x \in \mathcal{X}} \sigma^2(x).
\]

Therefore,
\[
\min_{\sigma^2(x) > 0, E[\tau^2(X_i) \sigma^2(x)] = 1} E[\tau^2(X_i) \sigma^2(X_i)] \geq \min_{x \in \mathcal{X}} \sigma^2(x).
\]

Suppose that $x_0 \in \mathcal{X}$ achieves the minimum of $\sigma^2(x)$ over $x \in \mathcal{X}$, that is,
\[
\sigma^2(x_0) \leq \sigma^2(x) \text{ for all } x \in \mathcal{X}.
\]

We consider two cases separately, (i) when $X_i$ is discrete and (ii) when $X_i$ is continuous.

(i) Suppose that the support of $X_i, \mathcal{X}$, is discrete. We set
\[
\sigma^2(x) = \begin{cases} \frac{1}{\tau^2(x_0) \sigma^2(x_0)} & \text{for } x = x_0 \\ 0 & \text{for } x \neq x_0. \end{cases}
\]

Then, we can see that $E[\tau^2(X_i) \sigma^2(X_i)] = 1$ and also
\[
E[\tau^2(X_i) \sigma^2(X_i) \sigma^2(X_i)] = \sigma^2(x_0).
\]

This implies that the lower bound of $\min_{\sigma^2(x) > 0, E[\tau^2(X_i) \sigma^2(x)] = 1}$ is achieved with
\[
\min_{\sigma^2(x) > 0, E[\tau^2(X_i) \sigma^2(x)] = 1} E[\tau^2(X_i) \sigma^2(X_i)] = \sigma^2(x_0)
\]
\[
= \min_{x \in \mathcal{X}} \sigma^2(x),
\]

and
\[
\min_{\sigma^2(x) > 0, E[\tau^2(X_i) \sigma^2(x)] = 1} \frac{E[\tau^2(X_i) \sigma^2(X_i)]}{E[\tau^2(X_i) \sigma^2(X_i)]} = \left( \sigma^2(x_0) \right)^2.
\]

(ii) When $X_i$ is a continuous random variable, we consider a sequence of functions $\sigma_{x,m}(x)$ such that
\[
\sigma_{x,m}(x) = \begin{cases} \left( f^{x_0 + \frac{1}{2m}}_{x_0 - \frac{1}{2m}} \frac{1}{\tau^2(x) p(x)} dx \right)^{-1} & \text{if } x_0 - \frac{1}{2m} < x < x_0 + \frac{1}{2m}, \\ 0 & \text{otherwise}. \end{cases}
\]

Then, for all $m$, we have
\[
E[\tau^2(X_i) \sigma_{x,m}(X_i)] = \left( \int_{x_0 - \frac{1}{2m}}^{x_0 + \frac{1}{2m}} \frac{1}{\tau^2(x) p(x)} dx \right)^{-1} \int \tau^2(x) p(x) dx \\
= \sigma^2(x_0).
\]

Also, as $m \to \infty$, we have
\[
E[\tau^2(X_i) \sigma^2(X_i)] \sigma_{x,m}(X_i)] = m \int \tau^2(x) \sigma^2(x) p(x) \left\{ x_0 - \frac{1}{2m} < x < x_0 + \frac{1}{2m} \right\} \frac{1}{\tau^2(x) p(x)} dx \\
\to \sigma^2(x_0)
\]

as in (A.8).

Now the econometrician moves. He wants to maximize the power by maximizing this ratio
\[
\frac{\left( \sigma^2(x_0) \right)^2}{E[\tau^2(X_i) \sigma^2(X_i)]^2}.
\]

manipulating $\sigma^2(\cdot)$. For this purpose, we can see that
\[
E[\tau^2(X_i) \sigma^2(X_i)]^2 \geq E[\tau^2(X_i) \inf_{x \in \mathcal{X}} \sigma^2(\cdot)]^2 = \sigma^4(x_0) E[\tau^2(X_i)]
\]

and hence,
\[
\frac{\left( \sigma^2(x_0) \right)^2}{E[\tau^2(X_i) \sigma^2(X_i)]^2} \leq \frac{\left( \sigma^4(x_0) \right)}{E[\tau^2(X_i)]^2} = E[\tau^2(X_i)]^{-1}.
\]

Note that the upper bound of $\frac{\left( \sigma^4(x_0) \right)}{E[\tau^2(X_i) \sigma^2(X_i)]^2}$, $\frac{1}{E[\tau^2(X_i)]}$, is achievable by choosing
\[
\sigma^2(x) = \sigma^2(x_0).
\]

This justifies a constant choice for $\sigma^2(x)$ from a minimax decision perspective.

### APPENDIX B: MULTIVARIATE EXTENSION

#### B.1 Score

Now suppose that $\xi, w \in \mathbb{R}^k$, while $n$ is a scalar. We interpret $\xi + \sqrt{n}w$ is a random deviation from $\xi$ in the random direction $w$ with
magnitude $\sqrt{\eta}$. As before, the mixture density is

$$h(y|x, \xi, \eta) = \int f(y|x, \xi + \sqrt{\eta} w) g(w|x) dw.$$  

Notice that

$$\frac{\partial}{\partial \eta} f(y|x, \xi + \sqrt{\eta} w) = \frac{f_{y|x}(y|x, \xi + \sqrt{\eta} w) w}{2\sqrt{\eta}}$$

and

$$\frac{\partial}{\partial \eta} f_{y|x}(y|x, \xi + \sqrt{\eta} w) w = \frac{1}{2\sqrt{\eta}} f_{y|x}(y|x, \xi + \sqrt{\eta} w) w.$$  

Then, the score function is equal to

$$s_{\eta}(y, \xi, \eta) = \frac{1}{h(y|x, \xi, \eta)} \int \frac{\partial f(y|x, \xi + \sqrt{\eta} w)}{\partial \eta} g(w|x) dw$$

$$= \frac{1}{h(y|x, \xi, \eta)} \int f_{y|x}(y|x, \xi + \sqrt{\eta} w) w g(w|x) dw$$

$$= \frac{1}{2\sqrt{\eta}} \int f_{y|x}(y|x, \xi + \sqrt{\eta} w) w g(w|x) dw,$$

and under the null hypothesis that $\eta = 0$, the score becomes

$$\lim_{\eta \to 0} s_{\eta}(y, \xi, \eta)$$

$$= \frac{1}{2} \int f_{y|x}(y|x, \xi) w g(w|x) dw$$

where $tr$ is the trace operator.

Now we denote

$$\Sigma(x) = \int w w' g(w|x) dw.$$  

Then, the score function is

$$\lim_{\eta \to 0} s_{\eta}(y, \xi, \eta) = \frac{1}{2} \int f_{y|x}(y|x, \xi) tr \left( f_{y|x}(y|x, \xi) \Sigma(x) \right)$$

$$= \frac{1}{2} \text{vec}(\Sigma(x))' \frac{1}{f(y|x, \xi)} \text{vec}(f_{y|x}(y|x, \xi)),$$

where vec is the vectorization operator that stacks the columns of the matrix.

B.2 The Mean and Variance Under the Local Alternative

Recall that for a $(K \times K)$ symmetric matrix $A$, we define half-vectors that eliminate all the supradiagonal elements of $A$, vec$(A)$. Let $D_k$ denote the $K^2 \times \frac{1}{2}K(K+1)$ duplication matrix such that $D_k \text{vec}(A) = \text{vec}(A)$ and $\text{vec}(A) = D_k' \text{vec}(A)$, where $D_k' = (D_k D_k)^{-1} D_k$ is the Moore–Penrose inverse of $D_k$.

Define

$$\Gamma(X) = E \left[ \left( \frac{1}{2} f_{y|x}(y|X, \xi) \vech \left( f_{y|x}(y|X, \xi) \right) \right)^' \right.$$

$$\times \left. \left( \frac{1}{2} f_{y|x}(y|X, \xi) \vech \left( f_{y|x}(y|X, \xi) \right) \right) \right]|X.$$  

We assume that $\Gamma(X)$ is strictly positive definite,

$$\Gamma(X) > 0.$$  

By following similar argument in the previous section, we can show that under the local alternative $k(\cdot|\cdot, \eta) = f_{y|x}(y|x, \xi + \sqrt{\eta} w) g(w|x) dw$, the score test statistic

$$\int \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{vec}(\Sigma(X_i))' \frac{1}{f(Y_i|X_i, \xi)} \text{vec}(f_{y|x}(Y_i|X_i, \xi))$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^n \text{vec}(\Sigma(X_i))' D_k D_k \frac{1}{f(Y_i|X_i, \xi)}$$

$$\times \text{vec}(f_{y|x}(Y_i|X_i, \xi))$$

is asymptotically normal with mean equal to

$$E \left[ \left( \frac{1}{2} \text{vec}(\Sigma(X))' D_k D_k \frac{1}{f(Y|X, \xi)} \text{vec}(f_{y|x}(Y|X, \xi)) \right)^' \times \left( \frac{1}{2} f_{y|x}(Y|X, \xi) \text{vec}(f_{y|x}(Y|X, \xi))' D_k' D_k \text{vec}(\Sigma(X)) \right) \right]$$

and variance equal to

$$\Omega = E \left[ \text{vec}(\Sigma(X))' D_k D_k \Gamma(X) D_k' D_k \text{vec}(\Sigma(X)) \right].$$

The LM test statistics in this case is

$$T_{\alpha}(\Sigma)$$

$$= \frac{1}{n} \sum_{i=1}^n \text{vec}(\Sigma(X_i))' D_k D_k \Gamma(X_i) D_k' D_k \text{vec}(\Sigma(X_i))^{-1}$$

$$\times \frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{1}{2} f_{y|x}(Y_i|X_i, \xi) \text{vec}(f_{y|x}(Y_i|X_i, \xi))' D_k' D_k \text{vec}(\Sigma(X))$$

$$\times \text{vec}(\Sigma(X_i))].$$

B.3 Max-Min Property and Practical Choice of the LM Test

To derive the max–min property of the multivariate LM test $T_{\alpha}(\Sigma)$ in (B.1), we consider a special case when the local deviation parameters are not correlated and its conditional variance is the same,

$$\int w w' g(w|x) dw = \sigma^2(x) I_K.$$  

In this case, $\Sigma(X) = \sigma^2(x) I_K$ and $\Sigma_n(X) = \sigma^2_n(x) I_K$. Define

$$\Sigma(X) = \sigma^2(X) I_K$$

and let

$$\tau^2(X) = 1_K' \Sigma(X) 1_K,$$

where $1_K$ is the $K$-vector of ones. Then, the noncentrality parameter of the LM test statistic $T_{\alpha}$ under the local alternative is

$$\phi(\sigma^2(\cdot), \sigma^2_n(\cdot)) = \frac{E \left[ \tau^2(X) \sigma^2(X) \sigma^2_n(X) \right]}{E \left[ \tau^2(X) \sigma^2(X) \right]}.$$
and as we shown in Theorem 1, the max-min solution is a constant function of $\sigma^2(\cdot)$. This justifies the use of the following LM test in the multivariate case:

$$T_n = \left( \frac{1}{n} \sum_{i=1}^n Y_i \Xi(X_i) I_K \right)^{-1} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^n \sum_{k=1}^K 1 \frac{f_{X_i \gamma_k}}{2} f(Y_i | X_i, \xi) \right),$$

where $\Xi(X)$ is defined in (B.2).

**SUPPLEMENTARY MATERIALS**

The supplementary material contains some details of the derivation for Example 3, as well as an estimator of the asymptotic variance of the test statistic.

**ACKNOWLEDGMENTS**

Moon acknowledges financial supports from the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014S1A5A8012177).

[Received May 2014. Revised June 2015.]

**REFERENCES**

Ackerberg, D., Chen, X., and Hahn, J. (2012), “A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators,” *Review of Economics and Statistics*, 94, 481–498. [364]

Aguirregabiria, V., and Mira, P. (2002), “Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models,” *Econometrica*, 70, 1519–1543. [360]

——— (2007), “Sequential Estimation of Dynamic Discrete Games,” *Econometrica*, 75, 1–53. [360,363]

Aguirregabiria, V., and Nevo, A. (2013), “Recent Developments in Empirical IO: Dynamic Demand and Dynamic Games,” in *Advances in Economics and Econometrics*, Tenth World Congress (Econometric Society Monographs, Vol. 3), eds. D. Acemoglu, M. Arellano, and E. Deckel, New York: Cambridge University Press, pp. 53–122. [360]

Bajari, P., Benkard, C. L., and Levin, J. (2007), “Estimating Dynamic Models of Imperfect Competition,” *Econometrica*, 75, 1331–1370. [360,363]

Bajari, P., Hahn, J., Hong, H., and Ridder, G. (2011), “A Note on Semiparametric Estimation of Finite Mixtures of Discrete Choice Models with Application to Game Theoretic Models,” *International Economic Review*, 52, 807–824. [360,363]

Bajari, P., Hong, H., Krainer, J., and Nekipelov, D. (2010), “Estimating Static Models of Strategic Interactions,” *Journal of Business & Economic Statistics*, 28, 469–482. [363]

Balestra, P., and Nerlove, M. (1966), “Pooling Cross Section and Time Series Data in the Estimation of a Dynamic Model: The Demand for Natural Gas,” *Econometrica*, 34, 585–612. [359]

Bierens, H. J. (1990), “A Consistent Conditional Moment Test of Functional Form,” *Econometrica*, 58, 1443–1458. [361]

Bjorn, P. A., and Vuong, Q. (1984), “Simultaneous Equations Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Labor Force Participation” *SSWF* 537, California Institute of Technology. [363]

Breusch, T. S., and Pagan, A. R. (1980), “The Lagrange Multiplier Test and Its Application to Model Specification in Econometrics,” *Review of Economic Studies*, 47, 239–253. [359,360]

Card, D., and Giuliano, L. (2013), “Peer Effects and Multiple Equilibria in the Risky Behavior of Friend,” *Review of Economics and Statistics*, 95, 1130–1149. [363]

Chamberlain, G. (1984), “Panel Data,” in *Handbook of Econometrics* (Vol. II), eds. Z. Griliches, and M. D. Intriligator, Amsterdam: North Holland, pp. 1247–1318. [359]

Chen, X., Linton, O., and van Keilegom, I. (2003), “Estimation of Semiparametric Models when the Criterion Function is not Smooth,” *Econometrica*, 71, 1591–1608. [364]

Chesher, A. (1984), “Testing for Neglected Heterogeneity,” *Econometrica*, 52, 865–872. [359,360,361]

Cox, D. R. (1983), “Some Remarks on Overdispersion,” *Biometrika*, 70, 269–274. [359]

Donald, S., Imbens, G., and Newey, W. K. (2003), “Empirical Likelihood Estimation and Consistent Tests With Conditional Moment Restrictions,” *Journal of Econometrics*, 117, 55–93. [361]

Engle, R. F. (1984), “Wald, Likelihood Ratio, and Lagrange Multiplier Tests in Econometrics,” in *Handbook of Econometrics* (Vol. II), eds. Z. Griliches, and M. D. Intriligator, Amsterdam: North Holland, pp. 775–826. [359]

Hahn, J., Newey, W. K., and Smith, R. J. (2014), “Neglected Heterogeneity in Moment Condition Models,” *Journal of Econometrics*, 178, 86–100. [359,360]

Hotz, V. J., and Miller, R. A. (1993), “Conditional Choice Probabilities and the Estimation of Dynamic Models,” *Review of Economic Studies*, 60, 497–529. [363]

Kasahara, H., and Shimotsu, K. (2009), “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices,” *Econometrica*, 77, 135–155. [364]

Lee, L., and Chesher, A. (1986), “Specification Testing when Score Test Statistics Are Identically Zero,” *Journal of Econometrics*, 31, 121–149. [359,360,361]

Maddala, G. S. (1971), “The Use of Variance Components Models in Pooling Cross Section and Time Series Data,” *Econometrica*, 39, 341–358. [359]

Newey, W. K. (1985), “Maximum Likelihood Specification Testing and Conditional Moment Tests,” *Econometrica*, 53, 1047–1070. [361]

——— (1994), “The Asymptotic Variance of Semiparametric Estimators,” *Econometrica*, 62, 1349–1382. [364]

Pakes, A., Ostrovsky, M., and Berry, S. (2007), “Simple Estimators for the Parameters of Discrete Dynamic Games, With Entry/Exit Examples,” *RAND Journal of Economics*, 38, 373–399. [360,363]

Pesendorfer, M., and Schmidt-Dengler, P. (2008), “Asymptotic Least Squares Estimators for Dynamic Games,” *Review of Economic Studies*, 75, 901–928. [360,363]

Seim, K. (2006), “An Empirical Model of Firm Entry With Endogenous Product- Type Choices,” *RAND Journal of Economics*, 37, 619–640. [363]

Sweeting, A. (2009), “The Strategic Timing of Radio Commercials: An Empirical Analysis Using Multiple Equilibria,” *RAND Journal of Economics*, 40, 710–742. [360,363,364]

White, H. (1982), “Maximum Likelihood Estimation of Misspecified Models,” *Econometrica*, 50, 1–25. [359]