Recombinant Estimation for Normal-Form Games, with Applications to Auctions and Bargaining

By

David Lucking-Reiley and Charles H. Mullin[*]
Vanderbilt University

November 1999

Abstract

In empirical analysis of economic games, researchers frequently wish to estimate quantities describing group outcomes, such as the expected revenue in an auction or the mean allocative efficiency in a market experiment. For such applications, we propose an improved statistical estimation technique called "recombinant estimation." The technique takes observations of the complete strategy of each player and recombines them to compute all the possible group outcomes which could have resulted from different matches of players. We calculate the improvement in efficiency of the recombinant estimator relative to the standard estimator, and show how to estimate standard errors for the recombinant estimator for use in hypothesis testing. We present an application to a two-player sealed-bid auction and a two-player ultimatum bargaining game. In these applications, the improved efficiency of our estimator is equivalent to an increase of between 40% and 200% in the sample size. We discuss how to design game experiments in order to be able to take full advantage of recombinant estimation. Finally, we discuss practical computational issues, showing how one can avoid combinatorial explosions of computing time while still yielding significantly improved efficiency of estimation.

Keywords: recombinant estimation, normal-form game, auctions, bargaining.

1. Introduction

Economists frequently study behavior in simultaneous-move games with multiple players, from auctions to Bertrand oligopoly to public-good provision. Recent interest in behavioral game theory has generated a wealth of experimental data on individuals' play in a wide variety of such games. Data on the strategies chosen by players can be used to make inferences about various quantities of economic interest, such as auction revenues, market prices, and levels of public-good provision. Often, experiments on multi-player games have each subject randomly and arbitrarily grouped with anonymous opponents. In such simultaneous-move games with arbitrary groupings, it seems reasonable to assume that each player's strategy is unaffected by the identities of her rivals. This assumption implies that the hypothetical game outcomes which would result from grouping a subject with players she did not actually face are just as valid as the game outcome resulting from the group of players she did play against. In this paper, we develop an estimation technique, which we call "recombinant estimation," designed to extract maximal information from observed strategies in order to estimate various properties of simultaneous-move games.

For concreteness, we present the example that originally motivated this paper. List and Lucking-Reiley (1999) performed an experiment to compare two different sealed-bid auction formats.[1] They conducted 15 identical two-person, uniform-price auctions for the same good, obtaining data from a total of 30 bidders, all of whom faced the identical bidding situation. They repeated this procedure with 30 additional bidders in 15 different Vickrey auctions. To compare the two auction formats, they conducted t-tests, which involve estimating both the mean and the standard error of the mean for each auction. Recall that with n independent observations, the standard deviation of the mean equals the standard deviation of the data divided by the square root of n. In order to test

---

[1] We return to this example in more detail in Section 4.

whether the mean bids were equal across auction formats, they computed the mean bid, and computed the standard error as the standard deviation divided by the square root of 30. They also compared revenues across auction formats, but since there were only 15 revenue observations (15 auctions with 2 bidders each), the standard error for this case was equal to the standard deviation divided only by the square root of 15. This seemed somehow unfair, as the same amount of information led to relatively higher standard errors for revenues than for bids.

This observation led us to realize that in a sealed-bid auction, the pairings of bidders were arbitrary, so we could observe the revenues that would have resulted from all 435 possible pairings (the number of combinations of 30 bidders taken two at a time). Computing the mean over these 435 different revenue observation clearly gives a more precise estimate of the mean revenue in the population of auctions with all possible bidders, since it uses more information. But what should be the standard error of the mean? Surely one can't just divide the standard deviation by the square root of 435; there's not *that* much information available! In this paper, we formalize the intuition just discussed, presenting a recombinant estimator for use in a variety of normal-form game applications, and deriving correct standard error estimates so that the estimator may be used for inference purposes.

Initially, we develop the implications of including these hypothetical pairings for two-player symmetric games in which the econometrician observes the complete strategy of each player. After illustrating the intuition for the results in this simplified context, we present an extension to *k*-player symmetric games. Next, we remove the assumption of symmetry, considering two-player and *k*-player asymmetric games. In asymmetric games, the possibilities for recombination are still present, but more limited. Finally, we present an estimator for the general case of a normal form game, which may include some players who are symmetric and other who are asymmetric with respect to each other.

We apply this new technique in two empirical examples. First, we compare Vickrey versus uniform-price sealed-bid auctions, using data from List and Lucking-Reiley (1999). The results show efficiency improvements of 28 to 68 percent under the new estimation technique, depending on the quantity estimated, which is roughly equivalent to increasing the sample size by 40 to 200 percent.

In our second application, we compute the expected profit to different strategies in a bargaining game, using data from Mitzkewitz and Nagel (1993). Although a bargaining game does not usually involve simultaneous moves, Mitzkewitz and Nagel use the "strategy method" to elicit complete contingent strategies, converting this extensive-form, sequential-move game into a normal-form, simultaneous-move game. The recombinant technique yields estimates 30 to 50 percent more efficient than the baseline (non-recombinant) estimator. Mitzkewitz and Nagel themselves computed recombinant estimates of expected profits, thus anticipating our work. We extend their results by putting their estimates into a general econometric framework, showing how to compute standard errors for these estimates, and computing the efficiency of their estimates relative to a baseline estimator.

We also consider the implications of our technique for the design of experiments. Some previous game experiments lend themselves easily to this estimation technique, while others do not. We discuss the classic auction experiment of Kagel, Harstad, and Levin (1987) as an example of the latter type, and show how the experiment could easily be redesigned in order to take full advantage of recombinant estimation. The key is to make sure that players face the same public environment in each session of the experiment, so that they may be recombined across sessions.

Finally, we discuss practical computational considerations. Recombinant estimation can involve a combinatorial explosion in the number of computations required. Depending on the number of players per group, a few dozen actual observations might spawn trillions of hypothetical observations to be computed. We show that one can

achieve almost all of the efficiency gains of recombinant estimation by considering only a random sample of several thousand of the hypothetical observations, an estimation task readily achieved by modern computers in seconds or less. Similar practical considerations apply to the estimation of standard errors for the estimator, where the combinatorial explosions are even more rapid than in the computation of the estimator itself.

Recombinant estimation bears some similarities to the statistical technique of bootstrapping, so it may be worth pointing out the differences. Bootstrapping provides a two-step procedure to estimate numerically the sampling distribution of an estimator whose distribution is impractical to compute analytically. The bootstrap first estimates the data-generating process from the empirical distribution of the data, and then uses Monte Carlo simulations of the data-generating process to estimate the sampling distribution of the estimator. Like bootstrapping, the technique involves reusing data in a method not practical before the advent of modern computing. But recombinant estimation constitutes a new, more efficient estimator for a population parameter, while bootstrapping is a method for doing inference with an existing estimator. Another contrast is that recombinant estimator does not involve simulated observations; it involves only the actual data observed by the econometrician. The crucial idea of the recombinant estimator is to consider the way individual strategies are combined to produce group outcomes, and to compute hypothetical outcomes which could have resulted from the observed strategies.

The remainder of the paper is organized as follows: Section 2 develops the estimator and its distribution. Section 3 provides two empirical examples of the estimator. Section 4 discusses how to design experiments in order to take advantage of this estimation technique. Section 5 presents practical computational considerations in the presence of an explosive number of recombinations. Finally, Section 6 concludes.

2. Estimators for Game Outcomes

In this section, we develop our recombinant estimator for the expected value of a group outcome in a simultaneous-move game. We start with relatively simple games, and add more generality in later subsections. In each case we consider, the estimator is always a sample average that satisfies the conditions of the central limit theorem, so it has an asymptotically normal sampling distribution.

Before developing the estimator in detail, we introduce some basic notation. Assume that the econometrician observes n groups of k players each, and let $m = nk$ denote the total number of players. Let x_i be the strategy observed for the ith player, with i taking on any integer value from 1 to m. Let p_j be the vector of players who participated in group j. For concreteness, we might think of the first group of players as $p_1 = (1,2,\ldots,k)$, the second group of players as $p_2 = (k+1, k+2, \ldots, 2k)$, and so on, up to the kth group of players $p_k = (m-k+1, m-k+2, \ldots, m)$. We can then write p_{jt} to represent the tth player in group j. Define y_j as the outcome which results from the strategies played by the set of players p_j, and let g be the function which maps strategies to group outcomes, so that $y_j = g\left(x_{p_{j1}}, x_{p_{j2}}, \ldots, x_{p_{jk}}\right)$. Note that both x_i and y_j might be vectors, but for ease of exposition, we assume y_j to be a scalar in this paper. Note further that the group outcome may be any function of the players' strategies which the econometrician might care to study. In particular, g is constrained to be neither differentiable nor continuous.

The researcher's goal is to estimate the expected value of y, the group outcome, in the entire population of possible players. Let y have a mean of μ and a variance of σ^2.[2] For comparison to the recombinant estimator, we first present the baseline estimator most commonly observed in practice. This baseline estimator is the sample average outcome

---

[2] Note that the outcome variable y is itself a random variable since it is a function of random variables (the strategies x_i).

across the *n* independent groups of players. That is, the baseline estimator involves computing the group outcome for each of the *n* groups, and computing the simple average of these *n* outcomes. We denote this baseline estimator by:

$$\hat{\mu} = \bar{y} = (1/n)\sum_{j=1}^{n} y_j \tag{2-1}$$

where *j* indexes the *n* groups. The variance of this estimator is

$$\text{var}(\bar{y}) = \sigma^2/n. \tag{2-2}$$

This estimator will be our reference point for efficiency comparisons to other proposed estimators.

In the following subsections, we first develop our recombinant estimator for symmetric two-player games, then generalize it to symmetric *k*-player games. We next develop a recombinant estimator for fully asymmetric 2-player games, and then generalize to asymmetric *k*-player games. The special cases provide a simplified setting in which to explain the intuition behind the recombinant estimation strategy. In the final subsection, we develop a general recombinant estimator for any *k*-player normal-form game, where each player may have symmetry with respect to some rivals but asymmetry with respect to others.

2.1 Symmetric Two-Player Games

Suppose we observe four players matched in two pairs in a two-player, symmetric game. Let y_1 represents the outcome for players 1 and 2, and y_2 represents the outcome for players 3 and 4. For convenience, we introduce the notation $y_{\{i,j\}}$ to represent the outcome resulting when players *i* and *j* play against each other, so that $y_1 = y_{\{1,2\}}$ and $y_2 = y_{\{3,4\}}$. Using the baseline estimator, the expected outcome would be estimated by the average of $y_{\{1,2\}}$ and $y_{\{3,4\}}$. However, one might be able to improve upon this estimator by considering alternative combinations of the players. In particular, the potential pairings {{1,3},{1,4},{2,3},{2,4}} and their associated outcomes are completely ignored

by the baseline estimator. The recombinant estimator that takes these potential pairings into account is:

$$\hat{\mu} = \left(\frac{2}{m(m-1)}\right) \sum_{i=1}^{m} \sum_{j=1}^{i-1} y_{\{i,j\}} . \qquad (2\text{-}3)$$

Note that the second summation runs from 1 to $i-1$ to avoid including both $y_{\{i,j\}}$ and $y_{\{j,i\}}$. Due to the symmetry in the outcome function these two outcomes correlate perfectly, so $y_{\{j,i\}}$ adds no new information once $y_{\{i,j\}}$ has been considered. There are $m(m-1)/2$ terms in the double summation; the above formula divides the sum by the number of terms in order to arrive at the sample average.

We have chosen to ignore the hypothetical outcomes which would result from matching a given player with herself. In some applications, the researcher may find it appropriate to include additional terms of the form $y_{\{i,i\}}$ in the estimator of the mean. The additional information would improve the efficiency of the estimator in small samples, but the effect is negligible in large samples, as the number of $y_{\{i,i\}}$ terms becomes negligible relative to the total number of $y_{\{i,j\}}$ terms. The two approaches are thus asymptotically equivalent, and we have chosen to ignore the $y_{\{i,i\}}$ terms for clarity of exposition. It is relatively straightforward, though notationally cumbersome, to adapt the equations in this paper to include these terms.

Some additional notation is required to present the variance of this estimator. Let $\varphi \equiv \text{cov}\left(y_{\{i,j\}}, y_{\{i,k\}}\right)$ for all $j \neq k$. Also, note that $\text{cov}\left(y_{\{i,j\}}, y_{\{k,l\}}\right) = 0$ when i, j, k and l are all unique. Then,

$$\text{var}(\hat{\mu}) = \left(\frac{2}{m(m-1)}\right)^2 \left[\sum_{i=1}^{m}\sum_{j=1}^{i-1}\text{var}\left(y_{\{i,j\}}\right) + \sum_{i=1}^{m}\sum_{j\neq i}\sum_{k\notin\{i,j\}}\text{cov}\left(y_{\{i,j\}}, y_{\{i,k\}}\right)\right]$$

$$= \left(\frac{2}{m(m-1)}\right)^2 \left[\frac{m(m-1)}{2}\sigma^2 + m(m-1)(m-2)\varphi\right]$$

$$= \frac{2\sigma^2}{m(m-1)} + \frac{4(m-2)\varphi}{m(m-1)}$$

$$\xrightarrow{asy} \frac{4\varphi}{m} = \frac{2\varphi}{n}$$

(2-4)

In general, for two pairs of players which share one player in common, we expect the covariance between the two outcomes to be positive. It is theoretically possible for the covariance to be zero, but negative values are ruled out by the symmetry of the outcome function g. For example, if bidder 1 in an auction bids relatively high, then an auction between bidders 1 and 2 and an auction between bidders 1 and 3 will both tend to have high revenues, and both will tend to have low revenues if bidder 1 bids relatively low.

From equation (2-4) it is clear that the variance of $\hat{\mu}$ converges at rate $\sqrt{n}$ to 2φ. Recall that the asymptotic variance of the baseline estimator converges at rate $\sqrt{n}$ to σ^2, so the relative efficiency gain is $\frac{2\psi}{\sigma^2}$. Proposition 1 in the following section proves that $2\varphi \leq \sigma^2$, so the recombinant estimator is never less efficient than the baseline estimator.

Note that for the special case where $\varphi = 0$ (which can happen for some distributions of strategies x together with some functional forms of the outcome function g), the recombinant estimator dominates the baseline estimator even more strongly. When $\varphi = 0$, the asymptotic variance of $\hat{\mu}$ converges at rate n to $\sigma^2/2$, while the baseline estimator converges only at rate $\sqrt{n}$.

For inference purposes, it is important to be able to estimate the variance of the estimator. To do this, one can use all the hypothetical outcomes $y_{\{i,j\}}$ to produce estimates of σ^2 and of φ, and then substitute these estimated values for the population

8

parameters in the variance formula above. We will present more details on computing estimated standard errors in section 6.2 of the paper.

2.2 Symmetric *K*-Player Games

With more than two players in a symmetric game, there are many more recombinations possible. Instead of merely $m(m-1)/2$ different hypothetical outcomes from the *m* players, we now have the number of combinations of *m* objects taken *k* at a time. Recall that p_j represents the set of players in group *j*. For the baseline estimator, *j* indexed only those groups actually observed, so it ranged only from 1 to *n*. Now we let P represent the collection of all possible *k*-player sets which can be taken from the *m* players, and let *j* index all the elements of P, so that *j* now ranges from 1 to $J \equiv m!/(k!(m-k)!)$. In this case, the estimator becomes:

$$\hat{\mu} = \left(\frac{1}{J}\right)\sum_{j=1}^{J} y_j \tag{2-5}$$

Note that the estimator in the previous subsection is a special case here: for $k = 2$, we have $J = m(m-1)/2$ terms in the summation.

To compute the asymptotic variance of this estimator, we again need to account for the fact that when two groups have at least one overlapping player, their outcomes are not independent observations, so there is some covariance between them. In principle, we might expect two groups with more than one overlapping player, such as {1,2,3} and {1,2,4} to have a higher outcome covariance than two groups with just a single overlapping player, such as {1,2,3} and {1,4,5}. However, it becomes cumbersome to keep track of the $k-1$ different types of covariances, and it turns out to be sufficient to consider the average covariance, denoted by φ, between any two groups that overlap by at least one player:

$$\varphi \equiv E\left\{\text{cov}(y_i, y_j) \middle| p_i \cap p_j \neq \varnothing \text{ and } p_i \neq p_j\right\} \tag{2-6}$$

where the expectation is taken with respect to the relative frequency with which each possible combination of repeated players occurs. Note that since each player's strategy is independent of all other players' strategies, when there is no intersection between two groups of players the covariance is zero.

The asymptotic variance of the estimator is:

$$
\begin{aligned}
\operatorname{var}(\hat{\mu}) &= (1/\mathrm{J})^2\left[\sum_{j=1}^{\mathrm{J}}\operatorname{var}(y_j)+\sum_{j=1}^{\mathrm{J}}\sum_{i\neq j}\operatorname{cov}(y_j,y_i)\right] \\
&= (1/\mathrm{J})^2\left[\mathrm{J}\,\sigma^2+\sum_{j=1}^{\mathrm{J}}\sum_{i\neq j}\operatorname{cov}(y_j,y_i)\right] \\
&\overset{asy}{\to}\frac{\sigma^2}{\left(m^k/k!\right)}+\frac{\left(m^{2k-1}\big/\left[(k-1)!(k-1)!\right]\right)}{\left(m^k/k!\right)^2}\varphi \\
&= \frac{\sigma^2}{\left(m^k/k!\right)}+\frac{k^2\varphi}{m} \\
&\overset{asy}{\to}\frac{k^2\varphi}{m}=\frac{k\varphi}{n}
\end{aligned}
\tag{2-7}
$$

The efficiency of this estimator relative to the baseline estimator is $k\varphi/\sigma^2$. Again, for the special case where $\varphi=0$, the recombinant estimator dominates the baseline estimator even more strongly, because it converges at rate $n^{k/2}$ while the baseline converges only at rate $n^{1/2}$.

2.3 Asymmetric Two-Player Games

In an asymmetric game, each rival player takes on a different role: buyer versus seller, bargain proposer versus responder, etc. In this case, there are fewer opportunities for recombination, because groupings are constrained always to take one player from each player type.[3] There are only $m/2$ choices for the first player in each pair, and $m/2$

---

[3] One way to increase the number of recombinations would be to put each player in both possible player roles, and elicit a player's strategy for both cases. This would in a sense restore symmetry, and allow many more recombinations. It is straightforward to adapt our estimator for this possibility, but because this

choices for the second player, for a total of only $J = m^2/4 = n^2$ possible pairs, by comparison $J = m(m-1)/2$ in the symmetric case.

Let S_1 be the set of individuals who played in the first player role, while S_2 is the set of individuals who played in the second player role. For concreteness, we can think of S_1 as the set of odd numbers no greater than m, and S_2 as the set of even numbers no greater than m. As before, let P represent the set of all possible pairs of players who could have been matched against each other: $P = \{(i,j)| i \in S_1 \text{ and } j \in S_2\}$. As noted above, there are n^2 elements in P. Letting p_j represent the jth element of P, the recombinant estimator can be written as:

$$\hat{\mu} = \left(\frac{1}{J}\right)\sum_{j=1}^{J} y_j . \tag{2-8}$$

To compute the variance of the recombinant estimator, we again need to consider covariances between terms which are not independent. For the asymmetric game, there are two different types of covariance to consider: covariance between outcomes which have the same first player, and covariance between outcomes which have the same second player:

(1) $\varphi_1 \equiv \text{cov}(y_i, y_j)$ for all $i \neq j$ such that $p_{i1} = p_{j1}$

(2) $\varphi_2 \equiv \text{cov}(y_i, y_j)$ for all $i \neq j$ such that $p_{i2} = p_{j2}$

Then the asymptotic variance of the estimator is:

$$\text{var}(\hat{\mu}) = \left(\frac{1}{J}\right)^2 \left[\sum_{j=1}^{J}\text{var}(y_j) + \sum_{j=1}^{J}\sum_{i \neq j}\text{cov}(y_j, y_i)\right]$$

$$= \left(\frac{1}{J}\right)^2 \left[J\sigma^2 + n^2(n-1)(\varphi_1 + \varphi_2)\right]$$

$$= \frac{\sigma^2}{n^2} + \frac{n^2(n-1)}{n^4}(\varphi_1 + \varphi_2)$$

$$\xrightarrow{asy} \frac{1}{n}(\varphi_1 + \varphi_2)$$

(2-9)

Thus, the efficiency gain relative to the baseline estimator is $\frac{(\varphi_1 + \varphi_2)}{\sigma^2}$. We will show in Proposition 1 that $\varphi_1 + \varphi_2 \leq \sigma^2$, so the recombinant estimator is never less efficient than the baseline estimator. Notice that if the game were symmetric, then $\varphi_1 = \varphi_2$ and the relative efficiency reduces to $\frac{2\varphi_1}{\sigma^2}$, exactly what we derived for the case of 2-player symmetric games.[4]

As before, in the special case $\varphi_1 = \varphi_2 = 0$, the recombinant estimator converges at rate n, while the baseline estimator converges at $\sqrt{n}$.

2.4 Fully Asymmetric *K*-Player Games

The generalization from 2–player asymmetric games to *k*-player asymmetric games is reasonably straightforward. We assume that the game is fully asymmetric, in the sense that each of the *k* different players enters the outcome function in a unique way. We will relax this assumption in section 2.5.

Generalizing the notation of the previous section, let S_i be the set of individuals who played in the *i*th player role. Then, the set of all possible groups of players becomes: $P = \left\{(i_1, \ldots, i_k) \mid i_j \in S_j \; \forall j \in \{1, \ldots, k\}\right\}$. There are $J = n^k$ elements in P. Letting p_j represent the *j*th element of P, the recombinant estimator can be written as:

---

[4] This result implies that restricting players to their original position in symmetric games, i.e. not allowing players who were initially in the same position to play against each other, does not reduce the asymptotic efficiency of the recombinant estimator. The reason for this result is illustrated in the section on computational methods.

$$\hat{\mu} = \left(\frac{1}{J}\right)\sum_{j=1}^{J} y_j .\tag{2-10}$$

The asymptotic variance of the estimator is:

$$\begin{aligned}
\mathrm{var}(\hat{\mu}) &= (1/J)^2\left[\sum_{j=1}^{J}\mathrm{var}(y_j) + \sum_{j=1}^{J}\sum_{i\neq j}\mathrm{cov}(y_j, y_i)\right] \\
&\xrightarrow{asy}\ \frac{\sigma^2}{n^k} + \frac{kn^{2k-1}}{n^{2k}}\varphi \\
&\xrightarrow{asy}\ \frac{k\varphi}{n}
\end{aligned}\tag{2-11}$$

The efficiency of this estimator relative to the baseline estimator is $k\varphi/\sigma^2$. Again, for the special case where $\varphi = 0$, the recombinant estimator converges at rate $n^{k/2}$ while the baseline converges only at rate $n^{1/2}$.

2.5 General K-Player Games

Some games of interest have both symmetry and asymmetry between players. For example, in a sealed-bid double-auction call market with b buyers and s sellers, the buyers are all symmetric to each other (they each have the same strategy space), and the sellers are all symmetric to each other, but there is asymmetry between buyers and sellers. Such a game has symmetry within player roles, but asymmetry between player roles.

Let t be the number of player roles and t_i be the number of players of type i in each group, e.g., $t = 2$, $t_1 = b$ and $t_2 = s$ in the double-auction example above. As before, let S_i be the set of individuals who played in the ith player role, e.g., S_1 contains all of the buyers in the above example. Then, the set of all possible groups of players becomes:

$$P = \left\{\left(i_{11},\ldots,i_{1t_1}\right),\ldots,\left(i_{t1},\ldots,i_{tt_t}\right)\middle| i_{j*} \in S_j\ \forall j \in \{1,\ldots,t\}\right\}.$$

Let J index the elements of P and p_j represent the jth element of P. Then, the recombinant estimator can be written as:

$$\hat{\mu} = \left(\frac{1}{J}\right)\sum_{j=1}^{J} y_j . \tag{2-12}$$

The asymptotic variance of the estimator is:

$$\mathrm{var}(\hat{\mu}) = (1/J)^2 \left[\sum_{j=1}^{J} \mathrm{var}(y_j) + \sum_{j=1}^{J}\sum_{i \neq j} \mathrm{cov}(y_j, y_i) \right]$$

$$\overset{asy}{\to} \frac{\sigma^2}{J^2} + \frac{J^2(k/n)}{J^2}\varphi \tag{2-13}$$

$$\overset{asy}{\to} \frac{k\varphi}{n}$$

The efficiency of this estimator relative to the baseline estimator is $k\varphi/\sigma^2$. Again, for the special case where $\varphi = 0$, the recombinant estimator converges at rate $n^{k/2}$ while the baseline converges only at rate $n^{1/2}$.

3. Theoretical Results on Recombinant Estimation

In this section, we present two propositions. The first proposition computes an upper bound on the size of the covariance term φ in k-player games. A corollary to this proposition is that the recombinant estimator is always at least as efficient as the baseline estimator. The second proposition describes under what circumstances the recombinant estimator's efficiency is greater than, rather than equal to, the efficiency of the baseline estimator.

Proposition 1 states that the average covariance between group outcomes which share overlapping strategies is less than one k^{th} of the variance of the group outcome function:

Proposition 1: $k\varphi \leq \mathrm{var}(y)$.

Proof: See Appendix.

An immediate corollary of Proposition 1 is that the recombinant estimator is always at least as efficient as the baseline estimator:

14

Corollary 1: The relative efficiency of the recombinant estimator, $k\varphi/\mathrm{var}(y)$, is less than or equal to one.

Proof: A direct application of Proposition 1.

Furthermore, the proof of Proposition 1 illustrates under what conditions the recombinant estimator is strictly more efficient than the baseline estimator. In particular, it is sufficient that the outcome function is not an additively separable function of the strategies, i.e. $y_j = g\left(x_{p_{j1}},\ldots,x_{p_{jk}}\right) \neq g_1\left(x_{p_{j1}}\right) + \ldots + g_k\left(x_{p_{jk}}\right).$[5]

Proposition 2: $\mathrm{var}(\hat{\mu}) \leq \mathrm{var}(\bar{y})$ with equality if and only if the outcome function is additively separable in the strategies.

Proof: Additive separability is required by property one used in the proof of Proposition 1.

Proposition 2 indicates that recombinant estimation provides strict efficiency gains for many different group outcome functions. The only exception is where the outcome is a linearly separable function of the players' strategies. To illustrate an example of this exception, suppose one is interested in the overall efficiency (relative to the Pareto optimum) which can be obtained in a public-good-contribution game. The game might involve k players each endowed with 100 units of money, which they can either keep or contribute to the public good. When player i contributes x_i to the public good, that money gets doubled and then divided equally among the players for their payoffs. In this game, the x_is are the strategies, and the group outcome is the total payoff to all players. The total payoff is equal to the total amount of public good provided plus the total amount of private good kept by the players, which turns out to be $100 + \sum_{i=1} x_i$. Since this is linear in the strategies, this game satisfies the conditions of Proposition 2, so

---

[5] If complete strategy information is gathered to allow for recombination across player roles, then asymmetry across player roles in the playoff function is also sufficient for the recombinant estimator to be strictly more efficient, i.e. g_i not equal to g_j for at least one pair of i and j.

there is no efficiency gain to be had. Intuitively, since the group outcome just involves summing up the individual outcomes, the recombinant estimator turns out to be numerically equivalent to the baseline estimator – it includes each term in the sum multiple times, but the average turns out to be exactly the same. For recombinant estimation to be useful, the group outcome of interest must not be linearly separable in the strategies. In the next section, we demonstrate two examples where the outcome function is nonlinear in the strategies, and compute the gains that occur. The first example is a symmetric game, while the second is an asymmetric game.

4. Applied Examples

We present two examples of recombinant estimation. First, we estimate expected revenues and the allocation of goods in multi-unit auctions, from experiments by List and Lucking-Reiley (1999). Second, we consider expected earnings of players in an ultimatum bargaining game, from experiments by Mitzkewitz and Nagel (1993).

4.1 Auctions

List and Lucking-Reiley (1999, henceforth LLR) compared two different sealed-bid auction experiments in an experiment at a sportscard trading show. Each of the 328 different bidders participated in a single two-bidder, two-unit auction for sportscards, with the items' book values ranging from $3 to $70. Half the bidders bid in auctions with the uniform-price highest-rejected-bid price rule, while the other half bid in auctions with the generalized Vickrey price rule. Each auction offered a pair of identical sportscards, and each bidder had the opportunity to submit two bids (one for each of the units of the good). After collecting the bid sheets from individuals over a period of several hours, the auctioneer randomly paired up individuals from the same type of auction in order to compute the results. Bidders were asked to return to the same booth at an appointed time in order to learn the results of the auction and conduct the transaction for any cards they had won.

LLR had four main results. First, second-unit bids were lower in the uniform-price treatment than in the Vickrey treatment, as predicted by Nash equilibrium theory. Second, first-unit bids were higher in the uniform-price treatment than in the Vickrey treatment, an effect unpredicted by theory. Third, the allocations of goods were significantly different across auction formats, with the uniform-price treatment resulting in more split allocations while the Vickrey treatment more frequently both cards went to the same bidder. Fourth, there was no statistically significant difference in revenues across auction formats. LLR's first two results involved analyses of the mean levels of bids. Since bids are individual strategies rather than group outcomes, recombinant estimation cannot improve econometric efficiency for mean bid levels. However, the last two results involve quantities (revenues and allocations) which are group outcomes. Since each bidder made his bids independently, we can use the rematching techniques of the present paper to provide improved estimates of the mean revenues and the proportion of split allocations for each auction format.

Tables 1 and 2 compare the baseline, nonrecombinant estimates presented by LLR with new, recombinant estimates of the same quantities. Table 1 presents results for the proportion of split allocations in each auction. There are results for ten different auction treatments: five Vickrey, and five uniform-price (the five different treatments involved different goods and different types of bidders, as shown in the rows of the table). Moving from the baseline to the recombinant estimator, the point estimates change by moderate amounts either up or down, and the standard errors uniformly become smaller. On average, the estimated variance of the recombinant estimator is 67.3 percent smaller than that of the baseline estimator. The efficiency improvement substantially increases the power of the t-test of the hypothesis that Vickrey and uniform-price auctions produce equivalent allocations; for example, the two-tailed p-value for this test in the first treatment (Sanders 1989, Nondealers) decreases from 0.04 with the baseline estimator to

0.0001 with the recombinant estimator. The improvement in efficiency is equivalent to more than a threefold increase in the sample size!

Table 2 presents similar results for auction revenues. Again, the standard errors generally decrease when moving from the baseline to the recombinant estimator, but in one of the ten treatments (Montana 1982, Dealers) the standard error actually increases with the recombinant estimator. As will be discussed in Section 6, this can happen because recombinant estimation produces a more accurate estimate of the estimator's variance in addition to the more accurate estimate of the mean. Thus, it is likely that the baseline estimator's estimated standard error was too small, especially given that the baseline estimate was based on a sample of only 10 auction revenue observations. Overall, the estimated variance of the recombinant estimator is 28.2 percent smaller than that of the baseline estimator. This improvement in efficiency is equivalent to about a 40 percent increase in the sample size.

4.2 Bargaining

Mitzkewitz and Nagel (1993) analyze the ultimatum game. Though a bargaining game does not typically involve simultaneous moves, Mitzkewitz and Nagel use the "strategy method" to collect data on complete contingent strategies, which converts the extensive-form, sequential-move bargaining game into a normal-form, simultaneous-move game. The basic idea of the ultimatum game is that Player A makes a take-it-or-leave-it offer to split a cash-valued "cake" with Player B. Player B either accepts the offer so that both players earn the indicated amounts, or rejects the offer and both players receive nothing. We consider the "demand form" of the game developed by Mitzkewitz and Nagel, in which the cake size is known to Player A but unknown to Player B. All Player B knows is the size of Player A's demand, not the size of the cake or the relative fraction A has demanded. The specific rules of the game are as follows:

1) The cake may be one of six sizes: 1, 2, 3, 4, 5, or 6 Taler, a fictitious currency worth approximately 70 cents.

2) Player A commits to an offer to player B for each of the six possible cake sizes. Offers are restricted to be an element of the set $\{0, 0.5, 1.0, \dots, 5.5, 6.0\}$ and cannot exceed the size of the cake.

3) Meanwhile, player B, with no knowledge of the cake size, decides which of the possible demands from zero to six she will accept, and which she will reject.

4) After both players have committed to their strategies, a die is rolled to determine the actual size of the cake. Given the cake size, player A's demand is revealed to player B, who either accepts or rejects it according to her previously elicited strategy. If accepted, player A receives what he demanded and player B receives the remainder of the cake. If rejected, both players receive zero.

This setting fits well the estimation technique of this paper. In fact, Mitzkewitz and Nagel themselves computed recombinant estimates of expected profits in their analysis (pg. 179). However, they did not provide standard errors of their estimates, nor did they evaluate the econometric efficiency of the procedure, so we extend their work in this direction. Table 3 contains the recombinant estimates,[6] their associated standard errors and the relative efficiency gain of recombinant estimator.[7]

---

[6] The point estimates presented in the table differ very slightly from those in Mitzkewitz and Nagel. The difference is attributable to the fact that Mitzkewitz and Nagel computed their estimates by hand.

[7] The experiment involved eight repeated trials, with each player facing a different rival each round (there were eight proposers and eight responders in each experimental session). Following Mizkewitz and Nagel, we recombine not just across individuals, but also across these rounds. It would be inappropriate to assume that the same individual's play is independent across different rounds, so we estimate our standard errors conservatively. We assume that there exists some positive covariance between observations involving the same player in different rounds, distinct from the positive covariance already assumed to exist between observations using the strategy of the same player in the same round. By estimating this new type of covariance from the data, we provide corrected (larger) standard errors which take into account this lack of independence. For ease of exposition, we omit an explicit discussion of the equations used for this purpose, which are a straightforward generalization of the equations already presented.

This estimation produces two interesting results. First, recombinant estimation yields a 16 and a 65 percent efficiency gains for the expected profits of players A and B, respectively. The greater gains are realized at the higher cake sizes (where the strategy space is relatively rich). These numbers, computed as in equation (2-9) above, indicate the decreased variance of the recombinant estimator relative to the baseline estimator (the simple average over only the realized pairings of players). Second, the estimated standard errors allow us to test an interesting hypothesis: whether the expected profit to player A is monotonically increasing in the cake size. Mitzkewitz and Nagel note that the estimated expected payoff to player A takes a maximum at cake size four and decreases from that point onward, so the point estimates look as if player A's profits are not monotonic in the cake size. However, they were unable to complete a formal test of this hypothesis, because they did not know the distribution of their estimates. The estimated expected revenue decreases by approximately 0.128 from a cake of size four to one of size six. The *t*-statistic for this change is 1.88, so we are able to reject at the 10% significance level the hypothesis that type-A players' profits are monotonically increasing in the cake size. Furthermore, the *t*-statistic associated with the baseline estimator is only 1.13, which would be insufficient evidence to reject the null hypothesis. Thus, in this example the recombinant estimation technique generates a formal rejection of the null hypothesis even though standard estimation techniques cannot do so.

5. Implications for Experimental Design

Our results show that recombinant estimation can produce drastic improvements in efficiency for estimating group outcomes from experimental data on auctions and other simultaneous-move games. There are also implications for the *design* of experiments. First, as noted above, recombinant estimation can be used anytime complete strategies are available. Therefore, experimenters who wish to take advantage of recombinant estimation in sequential games may wish to consider the use of the strategy method to

elicit complete contingent strategies, in order to have the right kind of data. Second, the experimenter needs to make sure that experimental conditions are held constant across experimental sessions in a way that allows for recombination. In the remainder of this section, we consider an example of an important experiment whose design could have been changed slightly along these lines, in order to generate data suitable for recombinant estimation.

Kagel, Harstad, and Levin (1987, henceforth KHL) test the theory of auctions with affiliated private values, with six bidders per auction, in a laboratory setting. Several of their experimental results involve estimates of nonlinear functions of all six bids; such estimates are prime candidates for improvement through our proposed technique. For example, they present a test of the key theoretical prediction that second-price sealed-bid auctions raise higher revenues than first-price auctions. In two of the three treatment conditions the difference is statistically insignificant, with high standard errors. Recombinant estimation could have reduced the variance of the estimated revenues in each auction type, improving the power of the hypothesis tests. Other examples in the paper where this estimation technique could have improved efficiency include: (1) comparing first-price auction revenues with and without public information, (2) measuring the effects of public information on efficiency, and (3) comparing average bidder profits when private values have high versus low variance.

In each experimental auction, KHL used the following procedure. First, they chose a locational parameter x_0 randomly from a uniform distribution on [\$25, \$125]. Second, they chose a set of six private values x_i for the six bidders, with each x_i drawn independently from the uniform distribution on $[x_0 - \varepsilon, x_0 + \varepsilon]$, where ε was a constant equal to either \$6, \$12, or \$24, depending on the round of the experiment. Each bidder learned his private value x_i, but did not necessarily learn the value of x_0, so he might remain uncertain about the distribution from which his value came. This generated "affiliation" between bidders' values, as described by Milgrom and Weber (1982). That

is, when a bidder learns that his value is high, he also learns that other bidders' values are likely to be relatively high.

After telling subjects their induced values x_i, KHL invited them to submit bids in an auction. Next they determined the winner of the auction, the price to be paid, and the profits earned by the winner, and informed the subjects of the results. They repeated this procedure for each round of the experiment, with 24 to 31 rounds per experimental session. Each session took place with a new group of six bidders, totaling seven sessions of first-price auctions and two sessions of second-price auctions. (KHL also report on sessions of English clock auctions, but their dynamic English auction cannot be represented as a normal-form game, so it is beyond the scope of this paper.)

KHL compute the difference between the observed first-price auction revenues and the theoretical revenues predicted by risk-neutral Nash equilibrium, and compute the mean of this value. (They compute the theoretical revenues separately for each realization of bidders' private values in the experiment.) They do the same thing for second-price auctions, and then compare the two means in a hypothesis test. Recombinant estimation might yield considerable improvement in the estimation of these means. Considering just a single round of the first-price auctions, the technique would involve taking the mean over 5.2 million different hypothetical revenue observations (combinations of 42 different subjects chosen six at a time), instead of the seven observed revenue observations. The exact amount of efficiency improvement would depend on the actual distribution of the data, but it is likely to be quite large.

The above example just considers a single round of the experiment. Recombinant estimation could be performed separately for each of the two dozen rounds of the experiment. One could even consider combining *across* rounds as well, with each bid in each round considered an independent draw from the population of bidding strategies. Of course, independence is a strong assumption, but it may be relaxed (as noted in footnote 6) to allow for correlation of strategies made by the same bidder in different rounds.

Recombination across rounds will sometimes not be possible at all; for example, one clearly would not want to combine across rounds with different values of ε. For their hypothesis tests KHL compute means both across groups and across rounds (only those rounds with the same values of ε). With 20 rounds of observations of 42 different bidders, the analogous recombinant estimator would involve *trillions* of different hypothetical revenue observations for each value of ε.

However, the locational parameter x_0 makes it impossible to recombine across rounds. Each different round of the experiment drew a different value of x_0 and bids submitted in response to different values of x_0 are not comparable to each other. For example, bids submitted when x_0=$40 could not be combined with bids submitted when x_0=$80, because the resulting revenues would be meaningless – such revenues could never actually result from a single auction. So separate estimation by round is required, which limits us to 5.2 million observations per round rather than trillions of observations overall. This is a fundamental restriction of the experiment; it would not be possible to study affiliated private values without having a location parameter x_0 that varied by round. (By contrast, experiments with independent private values, such as Cox, Roberson and Smith (1982) or Kagel and Levin (1993) would allow recombinations across rounds.)

Unfortunately, the specific experimental design of KHL does not even allow us to recombine across sessions within a single round. Because the value of the continuous random variable x_0 was drawn independently across experimental sessions as well as across rounds, there are no two groups of six bidders in the entire data set that can be recombined with each other. In order to use the recombinant estimator, one would have to redesign the experiment in the following way. Generate a set of random values of x_0 for the 26 rounds of the first experimental session, and then reuse these same values for each subsequent session. The values of x_0 remain random from the point of view of the subsequent sets of bidders, even though they are predetermined from the experimenter's point of view. Note that it is not necessary to use the same set of individual values x_i in

each experimental session, because the bids from bidders with different individual valuations could still be recombined, so long as the group-level parameters are the same. In fact, it is preferable to generate different random outcomes for the bidder values across sessions, because this would create a better representation of the full population to be studied: all bidders who might possibly have bid in an auction with a particular value of x_0. Only the group parameters need to be held constant across experimental sessions.

When would it be appropriate to recombine strategies across rounds as well as across independent groups of subjects? Note that the recombinant estimator assumes that strategies in different rounds are all draws from the same population. For example, in order to combine one individual's strategy in round 2 with another individual's strategy in round 10, one must believe that there is no systematic change in play over time. If, however, systematic learning about the game takes place over time (for example, bidders tend to gradually learn to bid lower), then one should not recombine across rounds, but only across sessions within the same round number. Another complication is the possibility of learning not just about the game but also about one's rival players: for example, one group of players learns to submit high bids because they played against a rival who bid unusually high in early rounds. If subjects do learn in this way about their specific group of rivals in repeated play, then any recombining across groups after the first round would be inappropriate.[8] In order to be able to make use of recombinant estimation in an experiment with repeated trials, the experimenter might wish to choose an experimental design which minimizes the possibility for learning about one's rivals. For example, Mitzkewitz and Nagel's experimental design ensured that over the eight rounds of the experiment, no player ever played against the same rival twice. Thus, in section 5.2, we feel relatively sanguine about recombining across rounds for that

_____

[8] It might be possible to construct a recombinant estimator which also controls for group effects of this type, but this is beyond the scope of the present paper.

experiment. An even stronger insurance would be the "zipper design" sometimes used in experiments, where in repeated rounds a player never comes into contact with anyone who might have been influenced by that player previously (that is, one never plays one's previous rivals again, nor one's rivals' rivals, nor one's rivals' rivals' rivals, and so on). In an experiment with repeated trials, minimizing players' repeated interactions with the same opponents provides a way to make better use of recombinant estimation.

To summarize, we find three key implications of our work for experimental design. First, recombinant estimation can be used in dynamic games if the experimenter can obtain complete contingent strategies. Second, in order to recombine across groups of players, all group-level parameters must be held constant for each group, though individual-level parameters may vary. Third, in order to use recombinant estimation in an experiment with repeated trials, it is important to have an experimental design which minimizes repeated contact between the same players.

6. Computational Methods

As discussed in the previous section, if KHL had been able to recombine across sessions, there would have been approximately 5.2 million combinations for each round. The growth of the number of combinations becomes ever more explosive as the number k of players per game increases. The same amount of data produces hundreds of combinations in a two-bidder auction and millions of combinations in a six-bidder auction. Even a relatively small number of observations (seven groups of six subjects each) can yield a very large number of recombinations. This exponential growth in the number of combinations could become computationally burdensome. Fortunately, it turns out that the vast majority of the efficiency gain from recombinant estimation can be obtained from a relatively small fraction of those possible recombinations. We show below that 100 combinations per player is usually sufficient. In KHL, this pares the full 5.2 million combinations per round down to a mere 4,200 combinations.

It turns out that this result is particularly important for computing estimated standard errors. Estimating standard errors requires an estimate of the covariance term phi, which itself can be computed via recombinant estimation with the "group outcome" appropriately defined. If the number of players per game is k, then the number of players per "group" in the estimation of φ is $2k-1$ (at least one player is in both groups), and this large number of players can produce enormous numbers of possible combinations when estimating standard errors. In practice, this turns out not to be a problem because a very accurate estimate of φ can be attained by considering a small subset of the possible terms . Because computation of standard errors is such an important concept, we give that topic its own subsection. First, however, we consider the simpler topic of the computation of point estimates.

6.1 Computationally Convenient Point Estimates

Another way to express the recombinant estimator is via iterated expectations. Recalling that P represents the set of all hypothetical groupings of the observed strategies, let P_i be the subset of P containing only those groups which include player i. That is, $P_i = \left\{ p_j \in P \mid i \in p_j \right\}$. Let J_i represent the number of elements in P_i. Now define $E\{y|i\}$ as the conditional expectation of y, restricted to groups containing player i. An estimate of this expectation is:

$$\hat{E}\{y|i\} = \left(\frac{1}{J_i}\right) \sum_{j:p_j \in P_i} y_j .$$
(6-1)

Then the recombinant estimator can be expressed as the sum of these estimated conditional expectation over all players i:

$$\hat{\mu} = \left(\frac{1}{m}\right) \sum_{i=1}^{m} \hat{E}\{y|i\} .^{9}$$
(6-2)

---

[9] Each $p_j \in P$ is in this summation k times. However, the sample average of J items is equal to the sample average of those same J items replicated k times.

In this section, we consider what happens if we estimate $\hat{E}\{y|i\}$ using less than the full set P_i of potential combinations involving player i. Suppose that we consider using r different combinations involving player i, where $r \leq J_i$. When $r = 1$, we obtain the baseline estimator, which involves no recombinations at all. Clearly, $r = 1$ gives a poor estimate of $\hat{E}\{y|i\}$. Including more recombinations allows the researcher to better estimate $\hat{E}\{y|i\}$ for each player. Let $\hat{E}_r\{y|i\}$ be the estimated expected value when r different combinations involving player i are considered, i.e.

$$\hat{E}_r\{y|i\} = \left(\frac{1}{r}\right) \sum_{j:p_j \in R_i} y_j \tag{6-3}$$

where R_i contains r random draws from P_i.

However, the gain to recombining is limited by the fact that all it provides is better estimates of $E\{y|i\}$, it does not increase the number of expected values to be considered. Therefore, the variance of the estimator is bounded below by the variance of

$$\hat{\mu} = \left(\frac{1}{m}\right) \sum_{i=1}^{m} E\{y|i\}. \tag{6-4}$$

Furthermore, the variance of the overall estimator is proportional to the $\text{var}(E\{y|i\})$ plus $\text{var}(\hat{E}\{y|i\}|i)$. Therefore, once the $\text{var}(\hat{E}\{y|i\}|i)$ is small relative to $\text{var}(E\{y|i\})$, there is not much to be gained by considering additional recombinations of the data.

Substituting in for $\hat{E}_r\{y|i\}$, the estimator can be expressed as

$$\hat{\mu} = \left(\frac{1}{m}\right) \sum_{i=1}^{m} \hat{E}_r\{y|i\} = \left(\frac{1}{m}\right) \sum_{i=1}^{m} \left(\frac{1}{r}\right) \sum_{j:p_j \in R_i} y_j. \tag{6-5}$$

The variance can be written as

$$\text{var}(\hat{\mu}) = \left(\frac{1}{m^2 r^2}\right) \left[mr \, \text{var}(y_j) + (m^2 r^2) E\{\text{cov}(y_j, y_i)\}\right]. \tag{6-6}$$

When y_j and y_i share at least one common player, the average $\text{cov}(y_j, y_i)$ is φ. The probability that p_j and p_i share at least one player is asymptotically equal to $(k^2 - 1/r)/m$ which is approximately equal to k^2/m. So,

27

$$\text{var}(\hat{\mu}) \approx \frac{\sigma^2}{mr} + \frac{k^2\varphi}{m} = \frac{\sigma^2}{knr} + \frac{k\varphi}{n}. \tag{6-7}$$

When r is equal to all possible combinations involving a particular player, then the first term in the variance is of lower order and we recover the variance that was derived earlier. However, if r is fixed, then the first term goes to zero at the same rate as the second, so we have a less efficient estimator. In particular, the relative efficiency of the recombinant estimator with r fixed with respect to the full blown recombinant estimator is

$$\left[\frac{\sigma^2/kr + k\varphi}{n}\right] \bigg/ \left[\frac{k\varphi}{n}\right] = 1 + \frac{\sigma^2}{rk^2\varphi}. \tag{6-8}$$

Alternatively, with respect to the baseline estimator the restricted alternative's relative efficiency is

$$\left[\frac{\sigma^2/kr + k\varphi}{n}\right] \bigg/ \left[\frac{\sigma^2}{n}\right] = \frac{k\varphi}{\sigma^2} + \frac{1}{kr} \tag{6-9}$$

which equals the relative efficiency of the recombinant estimator plus $1/kr$.

The exact gain from increasing the number of combinations considered is a function of the ratio between φ and σ^2. Table 4 displays the relative efficiencies of the baseline, recombinant and restricted estimators with r fixed at 50, 100 and 500. The table illustrates two points. First, as the covariance across recombinations with shared players decreases relative to the variance of the group outcome function, the potential gain from recombining data increases. Second, since the gain to recombining data is great when the covariance is small, the loss associated with restricting the estimator to consider r recombinations is greater when the covariance is small, as illustrated by the final column of the above table. In particular, consider the case where there are two players in the game and the covariance is 0.01. Here, restricting ourselves to 100 combinations results in a 25% efficiency loss relative to the recombinant estimator. However, we still capture 97.5% of the 98% efficiency gain relative to the baseline estimator.

6.2 Computationally Convenient Standard-Error Estimates

Computation of standard errors for the recombinant estimator requires separately estimating both σ^2 and φ, then substituting the results into the variance formula (2-13). Both σ^2 and φ can themselves be estimated via recombinant estimation, because they are themselves means of different group outcomes. To see this result, define a new outcome as the squared deviation from the average value of the old outcome, i.e.

$$s_j = \left(y_j - \hat{\mu}\right)^2. \tag{6-10}$$

Then the average value of this new outcome is the sample variance of the original outcome variable y, which makes it an estimate of the population variance σ^2. Since we have now posed the estimation of σ^2 as the estimation of a mean value, we can use the recombinant technique to estimate σ^2 more efficiently, computing this mean across all possible k-player groups. Equivalently, one generates a recombinant estimate of σ^2 by generating the same J different hypothetical outcomes $g(x)$ used to compute $\hat{\mu}$, and taking their sample standard deviation (whereas $\hat{\mu}$ was their sample mean).

Recombinant estimation can also be used to estimate φ. To see this, let $\Omega = \left\{p_i, p_j \in P \middle| p_i \cap p_j \neq \varnothing\right\}$. Define a group of players as $\omega_{ij} = \left\{p_i, p_j\right\} \in \Omega$ and the outcome as:

$$h_{ij} = \left(y_i - \hat{\mu}\right)\left(y_j - \hat{\mu}\right). \tag{6-11}$$

Then φ equals the expected value of this new outcome, restricted to observations where at least one player overlaps between groups i and j. Thus, φ is the expected value of a new type of group outcome, where this time the group size is $2k$-1 players (two different groups of size k which are constrained to overlap by at least one player). Since φ is the average value of a new type of group outcome, it can be estimated with the recombinant technique. In practice, to estimate φ one generates two columns of data, each row of which contains two different hypothetical outcomes which overlap by at least one player.

After generating all such possible pairs of outcomes, one takes the sample covariance between the two columns in order to estimate φ.

In general, recombinant estimation of φ is the most computationally expensive task we consider, because the number of such pairs of outcomes is of order n^{2k-1}, rather than the mere order n^k outcomes required for the recombinant estimator $\hat{\mu}$ or for the estimate of σ^2. Thus, in practical computation of standard errors, it can be even more essential to restrict attention to a subset of the possible combinations. Section 6.1 demonstrates that 100 recombinations per player is usually sufficient to obtain most of the gains from recombinant estimation. For practical purposes in estimating φ, we recommend generating at least 100 random pairs of outcomes overlapping in player 1, 100 pairs overlapping in player 2, and so on up to player m, then computing the sample covariance across all such pairs.

Because our estimate of σ^2 comes from recombinant estimation of an outcome (see equation (6-10)) which is nonlinear in the strategies, we expect this estimate to be strictly more efficient (see Proposition 2) than the standard non-recombinant estimate of σ^2 taken across only observed groups. If this result leads to more accurate estimates of the standard errors (recall that we have to estimate φ, as well as σ^2), then we will have even greater power in hypothesis testing. This result, although not fully developed here, is analogous to the power in a t-test. Recall that in a t-test, increasing the degrees of freedom (while holding the point estimate and the population standard deviation constant) also increases the power of the test. However, as with a t-test, the gain in power to increasing the precision of the estimate of the variance is decreasing in the precision of the original estimate. Thus, when the sample size is large, the recombinant and the simple estimators of the variance will produce tests of the same power (just as a t-test with 500 degrees of freedom has essentially the same power as one with 1000 degrees of freedom).

Finally, because the recombinant and the baseline estimators use different variance estimates, it is possible for the recombinant estimator to have a greater estimated

30

variance than the baseline estimator. This possibility is realized in one of the empirical examples (Table 2, Montana 1982, Dealers) in Section 4.

7. Concluding Remarks

We have presented a new econometric estimator for use in applications where the quantity of interest is a group outcome based upon the individual strategies of the players. Our recombinant estimator considers all possible hypothetical groupings, yielding considerable efficiency improvements relative to the standard estimator which considers only the observed groupings. In the applied examples we have examined so far, we find efficiency improvements equivalent to increases in the sample size of between 40 and 200 percent. Even greater efficiency improvements are possible in principle, but the exact amount of improvement depends on both the functional form of the group outcome function g, and on the distribution of the population of individual strategies. We have shown our estimator guarantees some efficiency improvement so long as g is not additively separable over the individual strategies x_i.[10]

This estimation technique would not have been feasible before the advent of modern computing, because calculating all possibly hypothetical outcomes requires considerable computing power. The applications we have considered, with tens of individual strategies and hundreds of possible groupings, have required less than a second to compute each estimate on a 400MHz computer. However, larger samples (with hundreds of individual strategy observations) can easily result in hundreds of trillions of combinations to compute, so complete computation of all possible combinations might overwhelm even today's computers. Fortunately, we have also shown that complete computation of all possible strategies is unnecessary; restricting attention to a

---

[10] Our intuition suggests that the efficiency improvements are greater in cases where $g(x)$ is more highly nonlinear and in cases where holding the total number of players fixed, the number of players per game is greater, though we offer no formal proof of either conjecture.

computationally manageable fraction of the possible combinations still yields almost all of the possible efficiency gains of recombinant estimation. This computational result is particularly useful for producing estimated standard errors for our estimator, which itself turns out to be an application of recombinant estimation where the number of possible combinations explodes very quickly.

Experiments on auctions and other normal-form games provide the most straightforward applications of our technique. Recombinant estimation can yield lower-variance estimates of such quantities as auction revenues, market efficiency, or the fraction of the time players achieve the Nash equilibrium in a matrix game. This technique may also help rescue data which might otherwise have become unusable when an experiment goes wrong. For example, suppose that in a twelve-person, simultaneous-move market experiment, one of the twelve participants leaves before submitting a strategy. The market efficiency for that group of twelve subjects cannot be calculated, because one can't find the actual allocation when one of the strategies is missing. However, the other eleven players all submitted valid strategies, so rather than throwing out these observations, one may recombine them with the data from other sessions in order to yield good market efficiency estimates.[11]

We have illustrated the recombinant-estimation technique in two applications. First, we have estimated revenues and allocations in two types of sealed-bid auction formats, pointing out the efficiency improvements relative to the standard estimator. We have also estimated the expected profits for each player in a bargaining game, and illustrated a hypothesis test where the null hypothesis cannot be rejected without the use of recombinant estimation. This latter example also illustrates how the technique may be extended to sequential games, if the experimenter is willing to convert the game from extensive form to normal form by eliciting complete contingent strategies from the

---

[11] We are grateful to Yan Chen for bringing this application to our attention.

players via the strategy method. As we have also shown in the context of auctions, there are other practical experimental design considerations which can make the difference between being able to use recombinant estimation and not being able to use it.

The estimator also holds some promise beyond simultaneous-move game experiments, and even beyond the field of economics. For example, consider a medical statistician who wishes to estimate the average ratio between healing time with drug A and healing time with drug B. Half the patients in the study received drug A, while the other half received drug B. This situation is isomorphic to the estimation of a two-player game, where the strategies are healing times and the outcome is the ratio between the healing times of one patient and another. The recombinant estimator considers all possible pairings of A-types with B-types, producing an estimate (and a standard error) for this ratio. The technique might be used in any statistical situation which involves matching or grouping of observations.

Future research might be able to generalize our results, for example by exploring the properties of recombinant estimation when the econometrician expects within-group fixed effects in repeated observations of the same group, or by finding additional applications for recombinant estimation. Most importantly, we hope that our work, with its discussions of the practical aspects of computation and of experimental design for use with this technique, will enable other researchers to improve the efficiency of their estimates of behavior in simultaneous-move games.

8. Appendix: Proofs of Propositions 1 and 2

This appendix contains a proof that the recombinant estimator is weakly asymptotically more efficient than the baseline estimator, i.e. Proposition 1. Furthermore, Proposition 2, that the recombinant estimator is strictly more efficient when the group outcome function is not additively separable in players' strategies, is proved as a special case. One lemma

and two properties of functions of independent random variables are invoked in the proof. Maintaining the notation of section 2, let

$$\varphi = E\left\{\text{cov}(y_i, y_j)\middle|\, p_i \cap p_j \neq \varnothing \text{ and } p_i \neq p_j\right\}.$$

The lemma states that the average covariance between groups with at least one common player converges to the average covariance between groups with exactly one common player. More precisely,

Lemma 1: $\lim\limits_{n\to\infty} \varphi = \psi$ where $\psi \equiv E\left\{\text{cov}(y_i, y_j)\middle|\, \exists \text{ exactly one pair } (t, \tau): p_{it} = p_{j\tau}\right\}$.

Proof: The number of pairs $\{y_i, y_j\}$ with exactly r repeated players is of order n^{2k-r}. Therefore, the proportion of the population of pairs with shared players that share exactly one player converges to one.

The two properties of functions of independent random variables are as follows:

1) Let $\mathbf{x}$ be a random k-dimensional vector where x_i is independent of x_j for all $i \neq j$. Then, for all $f(\mathbf{x}): R^k \to R$, f can be expressed as
   $$f(\mathbf{x}) = f_1(x_1) + \ldots + f_k(x_k) + \delta(\mathbf{x})$$
   where x_i is correlated with $f_i(x_i)$ and uncorrelated with all other terms in f. Furthermore,
   $$\sum_{i=1}^{k}\left[\text{corr}(f(\mathbf{x}), f_i(x_i))\right]^2 \leq 1$$
   with equality if and only if $\delta(\mathbf{x}) = 0$.

2) Let x be a random variable, $\mathbf{y}$ and $\mathbf{z}$ be k-dimensional random vectors where x, $\mathbf{y}$ and $\mathbf{z}$ are all independent of each other. Then, for all $f(\mathbf{x}): R^{k+1} \to R$
   $$\left[\text{corr}(f(x, \mathbf{y}), f(x, \mathbf{z}))\right]^2 = \left[\text{corr}(f(x, \mathbf{y}), f_x(x))\right]^2\left[\text{corr}(f(x, \mathbf{z}), f_x(x))\right]^2$$
   where $f_x(x)$ is defined as in property (1).

Armed with these three results, we are prepared to prove Proposition 1, that the recombinant estimator is weakly more efficiency than the standard estimator. In the proof, we point out where the lack of additive separability in the group outcome function leads to a strictly more efficient estimator, proving Proposition 2.

PROPOSITION 1: Let x_i be independently and identically distributed random variables and $y_j = g\left(x_{p_{j1}}, \ldots, x_{p_{jk}}\right): R^k \to R$. Then, φ is asymptotically less than or equal to one k^{th} the $\text{var}(y)$.

Proof: By lemma 1, it is sufficient to prove that the claim holds true when we restrict attention to pairs of outcomes which share exactly one player.

Recall that each player is only repeated in the role in which she is observed and that the game is symmetric within each player role. Therefore, without loss of generality, assume that if a player participates in both y_i and y_j, she is in the same position in both games. There are k player positions, so there are k equally likely types of terms that share exactly one player with y_i in this manner: y_j^t for $t = 1$ to k, where y_j^t indicates that a player from the tth position has been repeated in game j.[12] Note that y_j^t is independent of y_j^τ for all $t \neq \tau$ since the common player with group i differs and the remaining players are drawn at random.

Let $\rho_{ij,t} = \text{corr}\left(y_i, y_j^t\right)$. By Property (2),

$$\rho_{ij,t}^2 = \rho_{gg_t}^2 \rho_{gg_t}^2 = \rho_{gg_t}^4 .$$

Therefore, the sum of the correlations over the k player positions is

$$\sum_{t=1}^{k} \rho_{ij,t} = \sum_{t=1}^{k} \sqrt{\rho_{ij,t}^2} = \sum_{t=1}^{k} \rho_{gg_t}^2 \leq 1$$

where the final inequality is a direct application of Property (1).

Note that the final inequality above becomes strict if $y_j = g\left(x_{p_{j1}}, \ldots, x_{p_{jk}}\right)$ cannot be expressed as an additively separable function of the elements of $\mathbf{x}$, i.e. $\delta(\mathbf{x}) \neq 0$ in the decomposition in property (1). Furthermore, a strict inequality will carry through to a strict inequality in the statement of Proposition 1. This means a strictly more efficient estimator, and thus proves Proposition 2.

---

[12] Each type is equally likely because the set of players is restricted to unique combinations where the order of players within a player role does not matter, i.e. the set P in the notation of sub-section 2.5.

Returning to the proof of Proposition 1, algebraic manipulation of the above inequality yields the desired result, i.e.

$$1 \geq \sum_{t=1}^{k} \rho_{ij,t} = \sum_{t=1}^{k} \frac{\text{cov}\left(y_i, y_j \middle| p_{it} = p_{jt}\right)}{\sqrt{\text{var}(y_i)\text{var}(y_j)}} = \sum_{t=1}^{k} \frac{\text{cov}\left(y_i, y_j \middle| p_{it} = p_{jt}\right)}{\text{var}(y)}$$

$$\Rightarrow \sum_{t=1}^{k} \text{cov}\left(y_i, y_j \middle| p_{it} = p_{jt}\right) \leq \text{var}(y)$$

$$\Rightarrow \frac{1}{k} \sum_{t=1}^{k} \text{cov}\left(y_i, y_j \middle| p_{it} = p_{jt}\right) \leq \frac{\text{var}(y)}{k}$$

$$\Rightarrow \varphi \leq \frac{\text{var}(y)}{k}$$

Since $\lim_{n \to \infty} \varphi = \frac{1}{k} \sum_{t=1}^{k} \text{cov}\left(y_i, y_j \middle| p_{it} = p_{jt}\right)$ by lemma 1 and the fact that each type of covariance in the summation is equally likely.

References

Cox, James C., Bruce Roberson and Vernon L. Smith, "Theory and Behavior of Single Object Auctions," in Vernon L. Smith, ed., Research in Experimental Economics. Greenwich, CT: JAI Press (1982), pp. 1-43.

Kagel, John H., Ronald M. Harstad and Dan Levin, "Information Impact and Allocation Rules in Auctions with Affiliated Private Values: A Laboratory Study," *Econometrica*, 55 (November, 1987): 1275 – 1304.

Kagel, John H., and Dan Levin, "Independent Private Value Auctions: Bidder Behavior in First-, Second-, and Third-Price Auctions with Varying Numbers of Bidders," Economic Journal (July, 1993), 103 (419), pp. 868-879.

List, John and David Lucking-Reiley, "Demand Reduction in Multi-Unit Auctions: Evidence from a Sportscard Field Experiment," American Economic Review, forthcoming (2000). [Available for download at http://www.vanderbilt.edu/econ/reiley/cv.html]

Milgrom, Paul R. and Robert J. Weber, "A Theory of Auctions and Competitive Bidding," *Econometrica*, 50 (1982): 1089 – 1122.

Mitzkewitz, Michael and Rosemarie Nagel, "Experimental Results on Ultimatum Games with Incomplete Information," *International Journal of Game Theory* 22 (1993): 171 – 198.

Table 1
Proportion of Auctions with the Two Units Split Between the Two Bidders:
A Comparison of the Baseline and Recombinant Estimators
(Standard Errors in Parentheses)

Trading Card	Vickrey		Uniform	
	LLR	Recombinant	LLR	Recombinant
Barry Sanders 1989 Score; BV = $70	0.56	0.54	0.85	0.87
Nondealers	(0.12)	(0.07)	(0.08)	(0.05)
Cal Ripken 1982 Topps; BV = $70	0.23	0.31	0.87	0.89
Dealers	(0.11)	(0.05)	(0.09)	(0.05)
Jordan 1989 Hoops; BV=$3	0.32	0.35	0.54	0.52
Nondealers	(0.09)	(0.05)	(0.09)	(0.06)
Montana 1982 Topps; BV = $3	0.57	0.52	0.45	0.58
Dealers	(0.14)	(0.08)	(0.14)	(0.08)
Montana 1982 Topps; BV=$3	0.47	0.49	0.53	0.56
Nondealers	(0.11)	(0.07)	(0.12)	(0.08)

Table 2
Average Revenue from Vickrey and Uniform Sealed Bid Auctions:
A Comparison of the Baseline and Recombinant Estimators
(Standard Errors in Parentheses)

Trading Card	Vickrey		Uniform	
	LLR	Recombinant	LLR	Recombinant
Barry Sanders 1989 Score; BV = $70	52.06	49.42	48.71	44.90
Nondealers	(8.25)	(5.83)	(7.83)	(6.00)
Cal Ripken 1982 Topps; BV = $70	72.87	73.86	76.13	73.86
Dealers	(6.52)	(5.69)	(5.50)	(5.11)
Jordan 1989 Hoops; BV=$3	1.13	1.11	1.71	1.72
Nondealers	(0.30)	(0.22)	(0.27)	(0.25)
Montana 1982 Topps; BV = $3	2.37	2.18	2.13	2.12
Dealers	(0.42)	(0.28)	(0.34)	(0.36)
Montana 1982 Topps; BV=$3	0.66	0.66	0.83	0.88
Nondealers	(0.18)	(0.16)	(0.26)	(0.22)

Table 3
Mean Expected Payoffs for Demanders and Repliers in the Ultimatum Game
(Standard Errors in Parentheses)

	Cake					
	1	2	3	4	5	6
Player A						
Point Estimate	0.783	1.441	1.874	1.945	1.916	1.816
	(0.014)	(0.031)	(0.043)	(0.049)	(0.059)	(0.065)
Efficiency Gain	0.287	0.383	0.555	0.655	0.626	0.633
Player B						
Point Estimate	0.167	0.327	0.525	0.910	1.248	1.528
	(0.012)	(0.022)	(0.027)	(0.034)	(0.047)	(0.064)
Efficiency Gain	0.164	0.210	0.262	0.388	0.404	0.451

Table 4
Relative Efficiency of the Computationally Restricted Estimator

Number of Players	phi/sigma^2	Recombinant/ Baseline	Restricted/ Baseline	Restricted/ Recombinant
50 Recombinations				
2	0.250	0.500	0.510	1.020
2	0.100	0.200	0.210	1.050
2	0.010	0.020	0.030	1.500
5	0.100	0.500	0.504	1.008
5	0.010	0.050	0.054	1.080
100 Recombinations				
2	0.250	0.500	0.505	1.010
2	0.100	0.200	0.205	1.025
2	0.010	0.020	0.025	1.250
5	0.100	0.500	0.502	1.004
5	0.010	0.050	0.052	1.040
500 Recombinations				
2	0.250	0.500	0.501	1.002
2	0.100	0.200	0.201	1.005
2	0.010	0.020	0.021	1.050
5	0.100	0.500	0.500	1.001
5	0.010	0.050	0.050	1.008