Strategy and Sample Selection
A Strategic Selection Estimator*

Lucas Leemann†

Original version: 06/16/2012 – Revised version: 10/22/2013,
(All changes are indicated by the use of purple font)

ABSTRACT

The development and proliferation of strategic estimators has narrowed the gap between theoretical models and empirical testing. But despite recent contributions that extend the basic strategic estimator, researchers have continued to neglect a classic social science phenomenon: selection.

Compared to non-strategic estimators, strategic models are even more prone to selection effects. First, external shocks or omitted variables can lead to correlated errors. Second, because the systematic parts of actors' utilities usually overlap on certain key variables, the two sets of explanatory variables are correlated. As a result, both the systematic and the stochastic components can be correlated. However, given that the estimates for the first mover are computed based on the potentially biased predicted probabilities of the second actor, we also generate biased estimates for the first actor.

In applied work researchers neglect the potential shortcomings due to selection bias. This paper presents an alternative strategic estimator that takes selection into account and allows scholars to obtain consistent, unbiased, and efficient estimates in the presence of both selection and strategic action. I present a Monte Carlo analysis as well as a real world application to illustrate the superior performance of this estimator relative to the standard practice.

* I would like to thank Noah Buckley, Shigeo Hirano, Jeff Lax, Pierce O'Reilly, Curt Signorino, Neelan Sircar, and Greg Wawro for helpful comments and discussions. Sara Bjerg Moller and Simon Collrad-Wexler provided insights on the IR example. I also thank Thomas Plümper for the feedback he gave me as a discussant at the 2012 EPSA meeting. I am deeply indebted to two anonymous reviewers whose comments – which were thoughtful and very extensive – have significantly improved this manuscript. An earlier version of this paper received the Giancarlo Doria Paper Prize. I am very grateful to CUIT for giving me access to Hotfoot (high performance computing resource) to run the simulations. Companion software (for R) that implements the estimator proposed in this paper will be provided shortly on the author's webpage. All remaining errors are mine.

† Graduate Student, Department of Political Science, Columbia University; International Affairs Building; 420 W 118th Street; New York City.

1. EMPIRICAL MODELS BASED ON THEORETICAL MODELS

The motivation for strategic estimators can be found in the tension between the seemingly incompatible assumptions of formal and empirical models. When scholars want to empirically test formal models, they face an unsatisfying situation. Where they assume that the world is strategic in the first place (theoretical model), they assume conditional independence in the latter (statistical model). On the one hand, statistical methods assume conditional independence of the observations. On the other hand, hypotheses born out of a formal model often imply strategic dependence (Hall, 2003).

First, the assumed strategic nature of the data generating process leads to statistically *dependent* observations. To take into account the strategic nature, we need to rely on nested models. A second potential problem is that comparative statics in general rely on some form of (generalized) linear model which in general assumes monotonic effects (Lewis and Schultz, 2003). This assumption does not have to hold for data generating processes that are strategic in their nature (e.g., Signorino, 1999, 286-288).

In the past fifteen years there have been a large number of contributions proposing alternative estimators which address these problems. Most noteworthy are the contributions by Signorino (1999, 2002, 2003a, 2003b, 2008) that successfully adapted work by McKelvey and Palfrey (1995, 1996, 1998) from experimental settings to observational data. One problem is that these estimators are prone to selection bias. As in the Heckman sample selection model, correlated errors and correlated covariates lead to biased estimates (Heckman, 1979). Although Signorino (2002) proposed an estimator which can account for strategy and selection it has, thus far, no been applied. This paper presents an alternative and easy to use estimator which allows researchers to model strategic interactions without foregoing potential selection issues.

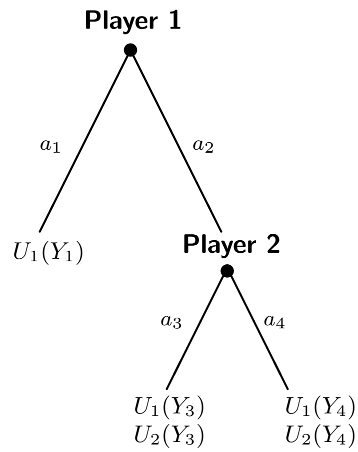
In this paper I start with a normal strategic estimator and show that it is prone to selection bias despite the fact that the selection stage is explicitly modeled. The reason for this is that the errors are constrained to be orthogonal and thereby induce selection bias. I derive an alternative version that does not suffer from sample selection bias. I show how this estimator differs from the one proposed by Signorino (2002) and is a novel addition to the empirical researchers toolbox.[1]

In the next two subsections I will lay out the basic theoretical model that underlies the observed behavior of actors and the corresponding statistical model (an agent error strategic estimator). In Section 2, I show why selection is a problem in general and how it is even more of a challenge for strategic estimators. Section 3 then shows the derivation of the proposed alternative estimator and how it deviates from known estimators. Finally, in Section 4 I show that the proposed estimator clearly outperforms the general strategic estimators which warrants the use of this improved strategic selection estimator.

1.1. *The Theoretical Model*

Many social science explanations rely on strategic actors who are assumed to anticipate each others' behavior and then choose an action such that their utility is maximized. We may consider a potential candidate in a Senate race trying to decide whether she or he should challenge the incumbent or not (Carson, 2005), a state deciding to go to war or not (Signorino and Tarar, 2006), an interest group deciding whether to launch a referendum or not (Hug and Leemann, 2011), or speculators deciding to attack currencies or not (Leblang, 2003). These models all share the basic feature that an actor has to decide which action to take but the final outcome is conditional on the actions that a second actor takes. In its simplest form such games can be theoretically represented by two actors, two nodes, two possible actions for each actor, and two information sets. In Figure 1 such a theoretical model is shown. Player 1 has two possible actions (a_1 and a_2), if Player 1 choses a_2, then Player 2 gets to decide among two options (a_3 and a_4). If Player 1 choses a_2 and Player 2 opts for a_3, we observe outcome 3. If Player 1 choses a_2 and Player 2 opts for a_4, we observe outcome 4. Each of the three different outcomes is associated with a specific utility to each actor

Figure 1: Theoretical Model



(the utility $U_2(Y_1)$ is omitted because it has no relevance for the equilibrium). If actors have full information and know each other's utilities, this game has a subgame perfect Nash equilibrium.

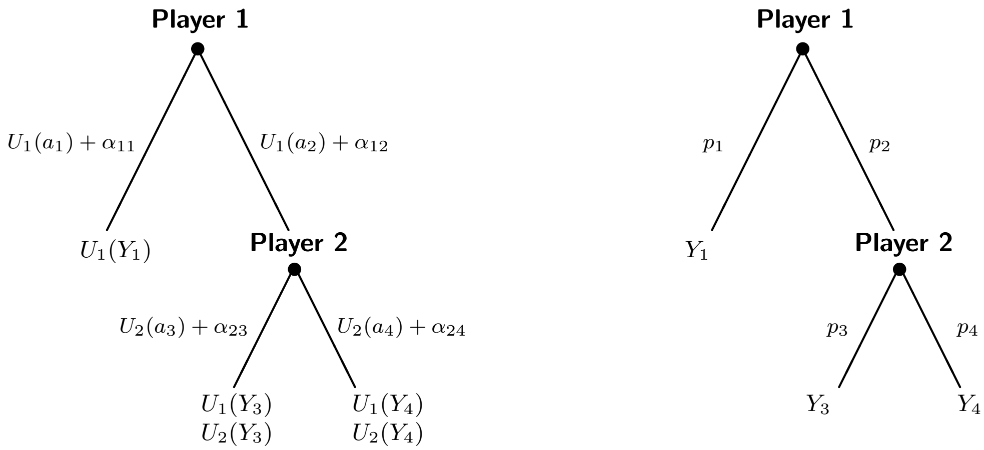
1.2. *Strategic Estimators*

If we have a theory as in Figure 1 and we want to test it on observed data there is a difficulty. Whereas the theoretical model assumes strategically dependent observations and non-monotonic effects, the empirical model assumes independence and usually monotonic effects. If one would apply a traditional empirical model it would be in its most simple form a logit or probit for Player 1 where the two outcomes would be whether Player 1 choses a_1 or a_2.

But the theoretical model is richer as it also entails strategic anticipation and that is not easily captured by adding covariates to a binary outcome model. The reason for this is that we would want to weight the expected probability for reaching outcome 3 (Player 2 choosing a_3) when Player 1 decides between a_1 and a_2. Signorino (1999) proposes to use the concept of Quantal Response Equilibrium Models and to apply it to observational data. This a paradigmatic example of empirical implications of theoretical models. If the theoretical model is that which is in Figure 1 one can formulate a statistical model that is exactly based upon the theoretical one.

To do so, we must have a source of error. While there are different possibilities for assigning sources of error, we will employ *agent error* at this point. Section 1.3 explores alternative sources of uncertainty. Agent error refers to a statistical model in which actors are assumed to misperceive the utility of their actions. It is assumed that the error pertains to the actions at the information set (Signorino, 2003*b*, 321-322).[2]

Figure 2: Empirical Model



We assume that the true utility for an actor is $U_m(a_k)^*$ where m denotes the actor and k the action. The true utility for Player 2 to chose action 3 is then $U_2(a_3)^* = U_2(a_3) + \alpha_{23}$, whereas $U_2(a_3)$ is the systematic part and α_{23} is the stochastic part. In what follows we use a random utility model to derive the estimator (Manski, 1977). Neither we as analysts nor the other actors are able to gauge the true utility for actor m from action k, but only a noisy measure (i.e. $U_m(a_k) \neq U_m(a_k)^*$).

The likelihood model is derived in terms of p_k. The log-likelihood can be expressed as follows when we assume that Y_1, Y_3, and Y_4 take on values 0 or 1:

$$\ell\ell \;=\; \sum_{i=1}^{n} \log \underbrace{(p_1)}_{=Pr(Y_1)} \cdot Y_{i1} + \log(\underbrace{p_{i2} \cdot p_{i3}}_{=Pr(Y_3)}) \cdot Y_{i3} + \log(\underbrace{p_{i2} \cdot p_{i4}}_{=Pr(Y_4)}) \cdot Y_{i4} \tag{1}$$

Probabilities p_3 and p_4 can be parametrized easily. For convenience, we will write this in terms of one random component:

$$\Delta\big(U_2(a_4, a_3)\big) \;=\; U_2(a_4) + \alpha_{24} - U_2(a_3) - \alpha_{23} \tag{2}$$
$$\Delta\big(U_2(a_4, a_3)\big) \;=\; U_2(a_4) - U_2(a_3) + \alpha_{24} - \alpha_{23} \tag{3}$$
$$\Delta\big(U_2(a_4, a_3)\big) \;=\; U_2(a_4) - U_2(a_3) + \varepsilon_2 \tag{4}$$

Instead of making assumptions regarding the two constituting error terms of the original formulation $(\alpha_{23}, \alpha_{24})$, we only make statements about their difference (ε_2).[3] Based on equation 4 we can formulate a model to estimate p_3 and p_4. To do so, we must impose a distributional assumption on ε_2. Here, we will assume $\varepsilon_2 \sim N(0, \sigma)$ and $\sigma = 1$ where the assumed variance is substantively meaningless and allows for identification. The Φ denotes the standard normal cumulative distribution. Hence:

$$p_4 \;=\; \Phi\left(\frac{U_2(a_4) - U_2(a_3)}{\sigma}\right) \tag{5}$$
$$p_4 \;=\; \Phi\left(\boldsymbol{\beta}_2' \mathbf{X}_2\right) \tag{6}$$

Equation 6 allows us to use observed data to express the difference in utility. To derive p_1 and p_2 is slightly more involved as the utility for a_2 depends on the actions of Player 2 (a_3, a_4) because the final outcome after Player 1 picks a_2 is determined by the action Player 2 takes when chosing a_3 or a_4. We use the same simplification as before and assume that $\alpha_{12} - \alpha_{11} = \varepsilon_1$ and that $\varepsilon_1 \sim N(0, \sigma)$ with $\sigma = 1$. Player 1 has to decide among two possible actions and the expected utility of a_2 is the weighted average of $U_1(Y_3)$ and $U_1(Y_4)$ whereas $E\big(U_1(a_2)\big) = p_3 \cdot U_1(Y_3) + p_4 \cdot U_1(Y_4)$.

$$p_2 = \Phi\left(\frac{U_1(a_2) - U_1(a_1)}{\sigma}\right) \tag{7}$$

$$p_2 = \Phi\left(\frac{p_3 \cdot U_1(Y_3) + p_4 \cdot U_1(Y_4) - U_1(Y_1)}{\sigma}\right) \tag{8}$$

$$p_2 = \Phi\left(p_3 \cdot \boldsymbol{\beta}'_{13}\mathbf{X}_{13} + p_4 \cdot \boldsymbol{\beta}'_{14}\mathbf{X}_{14} - \boldsymbol{\beta}'_{11}\mathbf{X}_{11}\right) \tag{9}$$

Since both actors have exactly two actions, we know that $p_1 = 1 - p_2$ and $p_3 = 1 - p_4$. Based on equations 6 and 9 we can formulate the log-likelihood as a function of data and parameters. This model can be estimated by FIML (full information maximum likelihood) or by a procedure called statistical backwards induction (Bas, Signorino, and Walker, 2008) where one uses a two-step procedure and relies on bootstrapping to correct second stage (first actor) estimates' uncertainties.

1.3. *Error Structure and Uncertainty*

It is worth adding two comments on the assumed uncertainty and the error structure implemented here. First, as noted above, instead of working with α_{23} and α_{24} we use ε_2. For agent error models this can be done without any loss of generality. But usually this is not done and the probabilities are presented as:

$$p_2 = \Phi\left(\frac{p_3 U_1(Y_3) + p_4 U_1(Y_4) - U_1(Y_1)}{\sqrt{\sigma^2_{\alpha_1} + \sigma^2_{\alpha_2}}}\right) \tag{10}$$

$$p_4 = \Phi\left(\frac{U_2(Y_4) - U_2(Y_3)}{\sqrt{\sigma^2_{\alpha_{23}} + \sigma^2_{\alpha_{24}}}}\right) \tag{11}$$

When implementing this model with statistical software, one has to set the variance of the random terms to achieve identification. Signorino and most authors usually set $\sigma_{\alpha_{ij}} = 1$ which leads to a denominator of $\sqrt{2}$.[4] In this paper, I assume that $\alpha_i - \alpha_j = \varepsilon$ and $\varepsilon \sim N(0, 1)$. Hence the denominator is 1 as usually with standard binary choice models (i.e. probit). It is important to note that it does not matter which value we assume as it is an arbitrary choice we have to make for identification. For what follows it is mathematically much more convenient to have one error term per actor than two.

Second, the error structure we assumed, agent error, is not only the closest to conventional models of discrete choice, but is also used frequently (i.e. Carson, 2003; Signorino, 2003; Gent, 2007). The prize-winning article deriving how strategic models can be estimated with standard software packages by Bas, Signorino, and Walker (2008) only considers agent error models. Finally, agent error models can be attractive since they are compatible with the notion of bounded rationality (Chen et al., 1997; Signorino, 2003*b*).

1.4. *Selection and Independence of Errors*

One of the crucial assumptions of the strategic estimator is the assumption that $cov(\varepsilon_1, \varepsilon_2) = 0$. It is assumed that the error of the first actor is uncorrelated with the error of the second actor. What is the error in a strategic model? There are at least three responses to that question. First, the error might be due to *regressor error* whereby the players act completely rational but the analyst cannot perfectly model their utilities with the explanatory variables, second, the error might be due to *agent error* which assumes that players may mis-perceive their utilities or make mistakes when implementing their strategies, and finally, the third source of error is *private information* whereby the players know more about their own utility than the other players and the analyst does (Signorino, 2003*b*). These three possible sources of the error represent ideal types but in any given application it is likely that the error stems from two or even all three sources. For simplicity however one assumes one type of error (Signorino, 2003*b*).

For straightforward strategic models one assumes that the errors are uncorrelated which enables us to derive Equation 1 as the product of independent events. To extend the statistical toolbox it seems natural to drop this assumption and derive an estimator which explicitly allows for correlated errors. One can assume that the correlated errors stem from common shocks (from the point of view of the analyst) and/or unmodeled variables which contribute to both players' utilities which the analyst does not incorporate to the model. In general, there are two ways to go about this. First, one can assume that the correlated errors are due to omitted variables and assume that the first player actually is aware of the full model of the second player. In statistical terms this amounts to assuming that the first player knows the correlation of his stochastic component with the stochastic component of the other player. Second, one can assume that the analyst and the first player share the same information with regard to the second player's utility, i.e. the first player does not know the correlation.

Why should the first player not know the correlation? One possibility is that the first player is not aware whether the second player also possesses this information. If state A is thinking about attacking state B, the public opinion in state B towards the current regime might be a relevant factor. If state A does not know whether the regime of state B is even aware of the public opinion it amounts to a situation in which one should assume that the correlation is unknown to Player 1. State B has an incentive to misrepresent the true utility resulting from its possible actions and to otherwise dissemble information (Fearon, 1995). Such a theoretical model is compatible with an agent error selection model which has not existed before now and is proposed in this paper.

In spite of these potential issues, in regular strategic models one assumes that the errors are uncorrelated. This is a strong assumption which is not necessary, as a strategic sample selection estimator is easily implemented. If this assumption does not hold, it is likely that strategic estimators will suffer from selection bias.[5] In this paper I present an alternative estimator which allows to incorporate the strategic nature of the theory and can retrieve unbiased estimates in the presence of sample selection.

2. SELECTION AND ITS CONSEQUENCES

Selection is a constant threat to valid empirical work in social science. A fundamental assumption made in most regression applications is that observations are conditionally independent. When analyzing data based on behavior of actors who self-select, the inferences may be biased. One form of selection may occur where the observability of observations depends on the outcome variable (i.e. truncation). Another form of selection might be that the sample is not a representative sample of the full population. To use an example from one of Heckman's original papers, the wages of migrants do not allow us to estimate what non-migrants would have earned (based on the wage function) if they had migrated (1979: 153). In the next two subsections I provide a brief review of the general issue of sample selection and the Heckman remedy as well as its shortcomings. I conclude Section 2 by illustrating the consequences of selection when the data generating process has a strategic nature.

2.1. *Classic Selection Bias*

We start by distinguishing between the actual regression function we are interested in and the selection stage model. The outcome of the latter determines whether an observation *selects* into the sample we use to estimate the actual regression function.

$$z_i^* = \boldsymbol{\gamma}'\mathbf{w}_i + u_i \quad (12)$$

$$y_i = \beta_0 + \boldsymbol{\beta}'\mathbf{x}_i + \epsilon_i \quad (13)$$

The sample that is available to estimate equation 13 (the regression function) upon depends on the z_i^*, such that observation i is in the sample if $z_i^* > 0$ and we do not observe case i (i.e. y_i is unknown) when $z_i^* \leq 0$. The latent variable z_i^* is modeled with the selection function in equation 12.

If we only have data for those observations which selected into the process we are estimating in equation 13 (those observations for which we have a measure for y_i) we might still be able to obtain unbiased estimates of $\hat{\beta}_0$ and $\boldsymbol{\beta}$. For that to be possible, one or both of these conditions must hold:

$$\text{cor}(u_i, \epsilon_i) = 0 \quad (14)$$

$$\text{cor}(\mathbf{w}_i, \mathbf{x}_i) = 0 \quad (15)$$

Only if the selection stage is unrelated to the regression stage, such that either the two random components are uncorrelated (equation 14) or there is no correlation between the systematic part of the selection stage and the regression stage (equation 15) will we recover unbiased estimates of $\boldsymbol{\beta}$ but not necessarily of β_0. If Equation 14 holds we can obtain unbiased estimates for β_0 as well, but if Equation 14 does not hold (while Equation 15 holds), one can obtain unbiased results of $\boldsymbol{\beta}$ but not for β_0. If the substantive interest lies in hypothesis testing rather than prediction, the bias in β_0 is not a problem (Achen, 1986, 78-79).[6] Note, that the bias does not stem from the fact that the distribution of $\mathbf{x}$ in the second stage sample, i.e. $f(\mathbf{x}|z* > 0)$, is different than the unconditional distribution of $\mathbf{x}$, i.e. $f(\mathbf{x})$. The bias stems from the fact that observations with low values on $\boldsymbol{\gamma}'\mathbf{w}_i$ $(= E(z_i^*))$ that get into the second stage sample must have a large random component. The large errors capture the effect of an unmeasured covariate and this large error causes the second stage error to be large as well (due to the correlation between u_i and ϵ_i). That in turn then leads to the *selection bias* in $\boldsymbol{\beta}$ (Sartori, 2003, 114).[7] Note that just estimating the regression model would lead to biased estimates of β_0 and $\boldsymbol{\beta}$ if $\rho \neq 0$ and if $\mathbf{w}_i$ and $\mathbf{x}_i$ are correlated. If only $\rho \neq 0$ the estimate for β_0 will be biased. The case in which the second stage regression is a binary model is analogous. If we are faced with two binary models where the first one models the selection of observations into the second one, All estimates will be biased if the random components and the systematic ones correlate (Equation 14 and Equation 15).

The challenge is to find a set of $\mathbf{w}_i$ that do not fully overlap with the $\mathbf{x}_i$ because the identification is then only driven by distributional assumptions (Sartori, 2003). For strategic models this is rarely a problem as a short look into applications of strategic estimators shows – usually there is some overlap but there remains a part of the systematic utility that is unique to each actor.

2.2. *Selection in Strategic Estimators: Theory and Practice*

In strategic estimators the situation deviates from the standard model presented in Heckman's (1974, 1976, 1979) work. The first difference is that the second stage model is itself binary and the second difference lies in the strategic anticipation of the next actor's move.

The dichotomous outcome in the second stage poses no greater problem. Boyes et al. (1989) derive an alternative to the Heckman selection estimator, in which the outcome regression is a binary model instead of a linear model. There a number of canned routines which are available in standard software (e.g. Stata: `heckprob`, R: `tobit2`). The second problem, the strategic nature, is more severe. If we have correlated errors of the two processes, the estimates will be biased. First, the second stage estimates, the estimates of Player 2, will be biased due to selection (at best only the intercept is biased). But in turn, the second stage estimates are used to generate weights for the two possible outcomes (Y_3, Y_4) and if these weights are biased, so will the estimates of Player 1's coefficients be biased.

All these potential shortcomings hinge on the assumption that the stochastic components (u_i, ϵ_i) are uncorrelated. As soon as they are correlated the results for Player 1 will be biased. In the normal selection model (as presented in Section 2.1) correlated errors alone will only bias the constant which means that we can still trust the estimates for the substantive variables ($\boldsymbol{\beta}$). But in the strategic models, the constant (β_0) feeds into the anticipated probability of Player 2's action and hence biases the first stage results. Whether the two errors are correlated or not is often hard to know *ex ante*.

From the bivariate selection model we already know that when the systematic parts ($\boldsymbol{\gamma}'\mathbf{w}_i$, $\boldsymbol{\beta}'\mathbf{x}_i$) are correlated as well as stochastic components, all estimates are potentially biased. In strategic models the systematic parts tend to be correlated because the first actor's utility contains the predicted probability of Player 2 choosing a_4 (to weigh the outcomes), the two systematic parts have to be correlated. In addition to this, many empirical examples display that a subset of variables is used to parameterize both the utilities of Player 1 and Player 2. Hence, if one should pay attention to sample selection bias in general, one should be on high alert when using strategic estimators. If the errors are correlated the results are biased.

Modeling Selection does not always take Selection into Account One point that might explain why the issue of selection is not taken more seriously in applications of strategic estimators may lie in the estimator itself. Since the strategic estimator models the selection stage (the first actor's decision) one might wrongly believe that this corrects for selection bias. But, as laid out in Section 2.1, the correction *works* through the correlated errors and when one assumes independent errors (error at the first node and the second node) one makes it impossible for the model to correct the second stage estimates. This is most clearly stated in Signorino (2002, 94) "*They are selection models because the actors select themselves and others into "subsamples" based on their choices. However, the strategic models implemented so far by Signorino (1999, 2000) differ from traditional bivariate* *probit selection models in that they assume independence of error terms across all decisions. This is by no means a necessary condition of strategic models. Rather, the assumption is made to simplify estimation.*"

Despite this clear explanation one can find many examples of scholars using strategic models and claiming that this automatically incorporates selection effects.[8] When using a strategic estimator one is modeling the selection stage (but by constraining the random components to be uncorrelated one hinders the extent to which the model corrects for selection bias, see Section 2.1).

Existing Solution In his 2002 paper, Signorino derives a strategic selection estimator that combines the strengths of both Heckman-type bivariate probit models as well as strategic estimators capable of handling strategic behavior data. The derived estimator is fairly complex as it is based on assumptions regarding the common knowledge of the error structure. In Figure 1 I show a model in which the assumption is that Player 1 knows the distribution of the random component of Player 2's utility ε_2 (or in Signorino's notation α_{23} and α_{24}). Signorino's selection estimator is thus built on the assumption that the *first actor is aware* of the correlation of the the two random components. Hence, when Player 1 assesses the expected utility from a_2 he or she will not have $E(\varepsilon_2|\varepsilon_1) = 0$ since the player knows his/her own draw (ε_1) and hence is aware of the correlation and will learn about Player 2's draw (ε_2) (Signorino, 2002, 104-105). But if one is not willing to make the assumption that Player 1 has more information than the analyst, one requires a statistical model which resembles an agent error model but allows for correlated disturbances. Note, that in applications of strategic estimators most analysts use an ordinary strategic model which rest on agent error and not on private information (e.g. Carson, 2003; Signorino, 2003; Gent, 2007).

Table 1: Overview of the relevant estimators

Abbreviation	Estimator	Anticipation	Errors ...	ρ known	Canned Procedure
Heck	**Heckman Selection estimator** (Boyes et al.,1989)	NO	correlated	NO	YES
Strat	**Strategic estimator** (Signorino,1999)	YES	uncorrelated	$\rho \equiv 0$	YES
SSc	**Strategic selection, common knowledge** (Signorino, 2002)	YES	correlated	YES	NO
SSa	**Strategic selection, agent error** (this paper)	YES	correlated	NO	YES

Notes: Anticipation: does Player 1 anticipate Player 2's action? Errors: Are the two errors correlated or assumed to be orthogonal? Canned Procedure: Is software readily available to estimate these models?

Estimation of the Signorino (2002) estimator is fairly complex. The implementation of the estimator relies on grid search to estimate the equilibrium probabilities. Based on the results of the grid search a standard maximizing function (here `maxlik` in Gauss) is used to estimate the parameters.[9] The complexity of the estimation makes it a less desirable estimator, a fact which is reflected by Signorino himself (2002, 106), who writes that "*[i]f the average user had a simpler and faster method, one that did not affect their inferences too much, that might be a reasonable and attractive option.*" (Signorino, 2002, 106). This seems to be the implicit message which many scholars took from Signorino (2002) – what has been subsequently adopted by researchers is unfortunately not a strategy that delivers unbiased estimates when facing selection. One has the impression that most readers gained the supposed insight that strategic estimators outperform bivariate selection models (Heckman-type bivariate probit).[10] While this is correct (strategic estimator (Strat) outperforms Heckman's non-strategic selection probit (Heck) if the DGP is strategic), it does not mean that one should choose the strategic estimator over a *strategic selection estimator (SSc)*.[11] I perform a literature review based on all 51 papers/book which cited Signorino's (2002) paper. Not a single of these articles actually employs the proposed strategic selection estimator. Hence, the standard is still to ignore any potential selection bias.

This mistaken impression, that the relevant decision regarding selection issues is whether one should use strategic models (Strat) or selection models (Heck), could have several sources. It might result from Signorino's claim that there are practical issues that make the strategic selection estimator (SSc) less appealing, or from the absence of direct comparisons of strategic and strategic selection estimators.[12] Either way, it seems that the problem that researchers have thought themselves to be facing is that the strategic selection estimator (SSc) was hard to use in practice, and that the relevant trade-off has appeared to be between the strategic estimator or a bivariate selection model (Heck, see footnote 10).

I set out first to derive an estimator that is as simple to implement as the standard strategic estimator but takes selection into account. I then perform a number of Monte Carlo simulations to illustrate its superiority over alternative possibilities. This is to my best knowledge the first time that a simulation study presents comparable results between a mis-specified strategic estimator and one that takes selection effects into account. In a second step I also change the data generating process and show that even if one employs the theoretical assumptions of Signorino (2002), the herein proposed estimator retrieves unbiased estimates. I believe that this serves to illustrate that scholars are no longer caught between a rock and a hard place.

3. A STRATEGIC SELECTION ESTIMATOR

To account for selection effects we need to have an estimator that does not require by assumption the independence of the two random components. It will facilitate the discussion to introduce some notation at this point. We can think of a strategic game as displayed in Figure 1 as a situation in which two actors are making two decisions, where Player 1 makes his or her choice in expectation of Player 2's choice. The unobserved latent variable is the difference in utility from the two options. Player 1's expected difference in utility from the two actions is denoted y_1^*, and for Player 2 it is y_2^*.

$$\Delta(U_1) = y_1^* \quad = \quad U_1(a_2) - U_1(a_1) + \varepsilon_1 \tag{16}$$
$$\Delta(U_2) = y_2^* \quad = \quad U_2(a_3) - U_2(a_4) + \varepsilon_2 \tag{17}$$

As detailed out for equations 6 to 9 we write ε_1 and ε_2 for the random components of this utility. As stated above (page 5) this notation is equivalent to the agent error model as described in detail in Signorino (2003*b*).

The derivation follows directly from the derivation of the non-strategic bivariate probit selection model. One can start out by recognizing how closely such an estimator is to the Heckman-type bivariate probit model (Boyes, Hoffman, and Low, 1989; Greene, 2003). We assume that we have three possible outcomes:

$$\begin{aligned}
a_2 = 0: \qquad\qquad P(Y_1) &= 1 - \Phi\left(\Delta U_1\right) && (18)\\
a_2 = 1, a3 = 1: P(Y_3) &= \mathbf{\Phi_2}\left(\Delta U_1; -\Delta U_2; -\rho\right) && (19)\\
a_2 = 1, a4 = 1: P(Y_4) &= \mathbf{\Phi_2}\left(\Delta U_1; \Delta U_2; \rho\right) && (20)
\end{aligned}$$

whereas Φ is the standard normal cumulative density and $\mathbf{\Phi_2}$ is the bivariate standard normal cumulative density with correlation ρ. We can also reformulate equations 18 to 20 in form of utility over outcomes, where $p_3 = 1 - p_4$. This is the simplified strategic selection estimator (SSa):

$$\begin{aligned}
a_2 = 0: \qquad\qquad P(Y_1) &= 1 - \Phi\big(p_3 U_1(Y_3) + p_4 U_1(Y_4) - U_1(Y_1)\big) && (21)\\
a_2 = 1, a3 = 1: P(Y_3) &= \mathbf{\Phi_2}\big(p_3 U_1(Y_3) + p_4 U_1(Y_4) - U_1(Y_1); -U_2(Y_4) + U_2(Y_3); -\rho\big) && (22)\\
a_2 = 1, a4 = 1: P(Y_4) &= \mathbf{\Phi_2}\big(p_3 U_1(Y_3) + p_4 U_1(Y_4) - U_1(Y_1); U_2(Y_4) - U_2(Y_3; \rho\big) && (23)\\
p_3 = 1 - p_4: \qquad\qquad p_4 &= \Phi\big(U_2(Y_4) - U_2(Y_3)\big) && (24)
\end{aligned}$$

The likelihood function, or its logarithm (log-likelihood), follow analogously from equation 1:

$$\begin{aligned}
\ell\ell &= \sum_{i=1}^{N} \log\left(\mathbf{\Phi_2}\left(\boldsymbol{\beta}_1'\mathbf{x}_{i1}, -\boldsymbol{\beta}_2'\mathbf{x}_{i2}, -\rho\right)^{(1-Y_{i1})(1-Y_{i4})}\right)\\
&\quad + \sum_{i=1}^{N} \log\left(\mathbf{\Phi_2}\left(\boldsymbol{\beta}_1'\mathbf{x}_{i1}, -\boldsymbol{\beta}_2'\mathbf{x}_{i2}, \rho\right)^{(1-Y_{i1})Y_{i4}}\right) + \sum_{i=1}^{N} \log\left(\left(1 - \Phi(\boldsymbol{\beta}_1'\mathbf{x}_{i1})\right)^{Y_{i1}}\right) \qquad (25)
\end{aligned}$$

The structural component enters when we parametrize the term $\Delta U_{1:a2,a1}$, the difference in utility for Player 1 from choosing action a_2 over a_1, with observable variables and estimable parameters. As in equation 9 it is assumed that Player 1 weighs the two possible subsequent outcomes Y_3 and Y_4 by the expected probability that Player 2 will chose a_3 or a_4. Unlike Signorino (2002) it is not necessary to impose the assumption that Player 1 knows the correlation of his or her error with the error of Player 2.[13] This estimator is hence based on weaker assumptions but the estimator by Signorino (2002) is not nested and the decision which estimator to use should be based on substantive assumptions regarding the knowledge of the actors. A potential data driven test could help to distinguish between models (Vuong, 1989; Clarke and Signorino, 2010). Note, that since we are simultaneously estimating the correlation between the random components the model provides a 'built-in' hypothesis test to whether there is selection bias or not.

Finally, as with the standard Heckman selection models there is an identification issue. Most strategic applications include a set of variables for Player 1 and a set of variables for Player 2 which do not fully overlap.[14] But there is a problem if the theoretical model requires the exact same set of explanatory variables for both players. The identification is then solely due to the distributional assumption as with the normal Heckman model (Heck). Sartori shows that this problem is especially prevalent if the sample size is small (Sartori, 2003).

4. TESTING PERFORMANCE: A MONTE CARLO SIMULATION

To test the performance of the strategic selection estimator (SSa) in comparison with some other commonly-used estimators (Strat, Heck), I performed Monte Carlo simulations. The Monte Carlo simulations are based on 20 experimental conditions. For each of these conditions 1000 simulations are performed. That means that there are 20,000 simulated data sets and on each of them three models are estimated. First, a strategic estimator with uncorrelated errors (Strat),[15] a bivariate selection model (Heckman probit, Heck), and a strategic selection model (SSa) as outlined in equations 21-24 are estimated. The simulations were run on a computer cluster and the files are available upon request.

4.1. *The Data Generating Process*

For each of the 20,000 simulations I create a sample of 1,000 observations and use this basic data generating process in which the utility for Player 1 from Y_1 is captured by $\beta_{11}x_1$, the utility for Y_3 is parametrized as constant $\beta_{10.2}$, and the utility from Y_4 is modeled as $\beta_{12}x_2$. Finally the outcome a_3 is modeled as the constant only. For Player 2 both x's are the utility for a_4 and a_3 is again modeled as the constant.

$$
\begin{aligned}
y_{1i} &= \beta_{10.1} + \beta_{11} \cdot x_{1i} + p4 \cdot \beta_{12} x_{2i} - p_3 \cdot \beta_{10.2} + \varepsilon_{1i} && (26)\\
y_{2i} &= \beta_{20} + \beta_{21} \cdot x_{3i} + \beta_{22} \cdot x_{2i} + \varepsilon_{2i} && (27)\\
\boldsymbol{\varepsilon} &\sim N_b(\mathbf{0}, \boldsymbol{\Sigma}) && (28)\\
\boldsymbol{\Sigma} &= \begin{bmatrix} \sigma^2_{\epsilon_1} & \rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2} \\ \rho\sigma_{\varepsilon_1}\sigma_{\varepsilon_2} & \sigma^2_{\varepsilon_2} \end{bmatrix} && (29)\\
x_k &\sim N(0, \sigma_k) \; \forall k = 1, 2, 3 && (30)
\end{aligned}
$$

The two random components ($\boldsymbol{\varepsilon}$) are assumed to stem from a bivariate normal distribution (N_b) with mean (0,0) and variance covariance matrix $\boldsymbol{\Sigma}$. Note, that the two relevant points in this data generating process are first that the random components may be correlated (eq. 29) and second that the systematic components are correlated since x_1 is part of Player 1's as well as Player 2's utility. Finally, to make statements based on the ratio of variation in the systematic part and the variation in the errors, one can pick σ_{ϵ_1}, σ_{ϵ_2}, and σ_k such that the systematic part of the utility has greater variance that then errors and vice versa by changing σ_{ϵ_1} and σ_{ϵ_2}. The coefficients are set to 1 but the constants are set to 0.

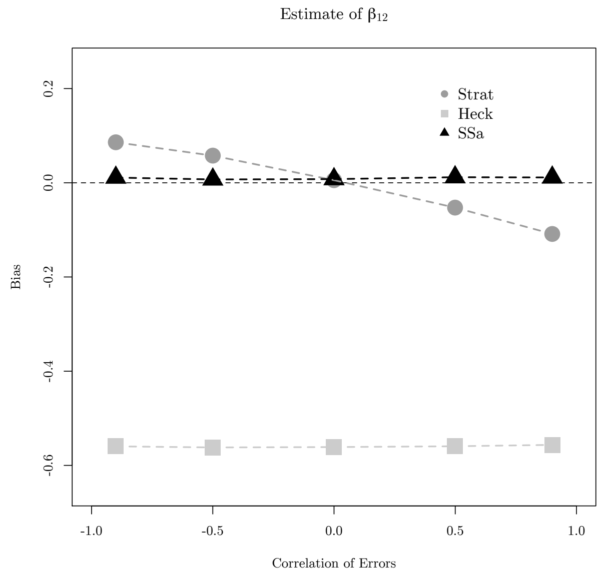
The 20 different experimental conditions result as the product of 5 different levels of correlation among the random components ($\rho = -0.9, -0.5, 0, 0.5, 0.9$) and four different ratios between the variance of the random and the systematic part ($(\sigma^2_{\epsilon_1} + \sigma^2_{\epsilon_2})/(2 \cdot \sigma^2_k) = 0.1, 0.5, 1.0, 2.0$) allowing me to mimic situations of varying signal to noise ratios.

4.2. *Results of the Monte Carlo Simulation*

Before presenting the results of the Monte Carlo simulation, we should recall the theoretical expectations we have. Most noteworthy are two points, first, that the strategic selection estimator (SSa) should outperform both the strategic estimator (Strat) as well as the selection estimator (Heck).

Second, that the strategic selection estimator (SSa) should outperform the strategic estimator (Strat) especially for estimates of $\boldsymbol{\beta}_2$ (the coefficients of the parametrized utility of Player 2). This is because in the strategic estimator (Strat) those estimates are identical to a pure binary model

Figure 3: Bias of Estimate, $\frac{\sigma^2_{\epsilon_1}+\sigma^2_{\epsilon_2}}{2\cdot\sigma^2_k}=1$



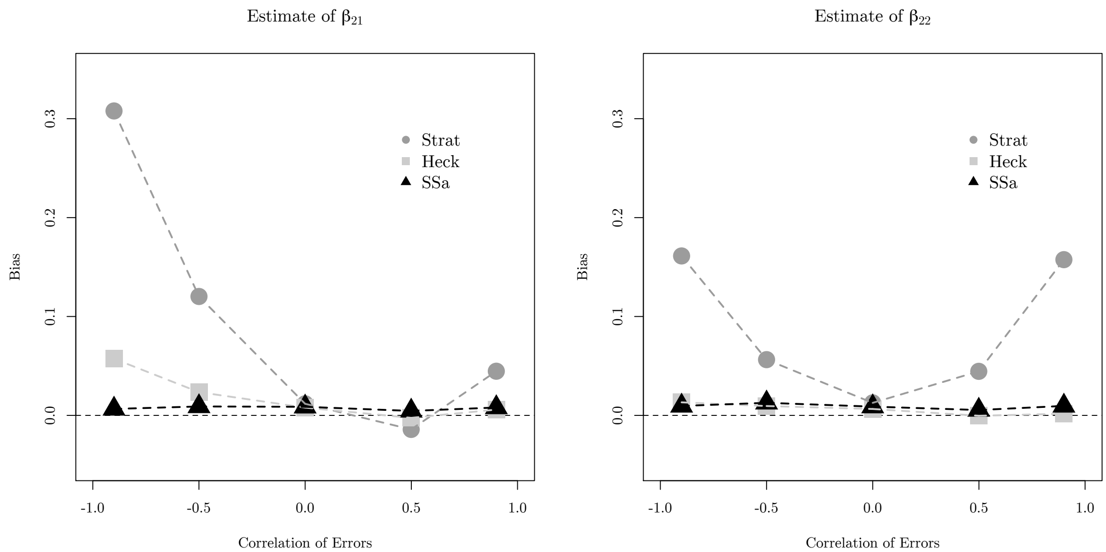
over the choice of the second actor based on those observations for which Player 1 chooses a_2 and one 'moves' down the game tree (the only information in the likelihood function for those estimates stem from the last two components of equation 1).[16]

Figure 3 shows all simulations for β_{12}. The $y-$axis shows the bias (bias= $E(\hat{\beta}_j - \beta_j)$). On the $x-$axis is the level of correlation between the two random components; the experimental conditions are $\rho \in \{-0.9, -0.5, 0, +0.5, +0.9\}$. The light grey squares represent the estimates obtained from the selection estimator (Heck), the grey circles are used for the estimates based on the strategic estimator (Strat), and black triangles show the estimates of the estimator derived earlier in this paper (strategic selection estimator, SSa). Each dot represents the average value for 1,000 simulations.

For parameter β_{12} there is a clear outcome; the selection estimator (Heck) – failing to accommodate the strategic nature of the DGP – performs poorly in comparison with the two estimators taking anticipation (strategic behavior) into account. As the the absolute value of the correlation between the two random components grows, the strategic selection estimator (SSa) outperforms the strategic estimator (Strat). This difference is not due to a direct effect of the selection bias. Since β_{12} is estimated in both models based on the anticipation of Player 2's move (hence: β_{20}, β_{21}, β_{22}), any bias in the latter three estimates will trickle upwards and bias β_{12}.

In Figure 4 two plots are displayed. The parameters associated with the utility of Player 2 are shown, whereas the left panel shows the performance for different estimators when estimating β_{21} and the right panel shows the results for β_{22}. A number of stylized facts can be learned from this figure. First, the strategic estimator (Strat), which does not take selection bias into account displays the strongest bias. Second, the selection estimator (Heck) is outperformed most significantly when the correlation of the errors is large and negative. Third, when the correlation is very close to 0, all three estimators perform close to identically. Finally, for β_{21} we see that the bias is not symmetric to the absolute value of ρ; especially if the correlation is negative, the strategic estimator (Strat) is more biased than if the correlation is positive. Altogether, these results seem to

Figure 4: Bias of Estimate, $\frac{\sigma^2_{\epsilon_1}+\sigma^2_{\epsilon_2}}{2\cdot\sigma^2_k} = 1$



be in stark difference to practice and also general advice given for strategic models with correlated errors (Signorino, 2002).

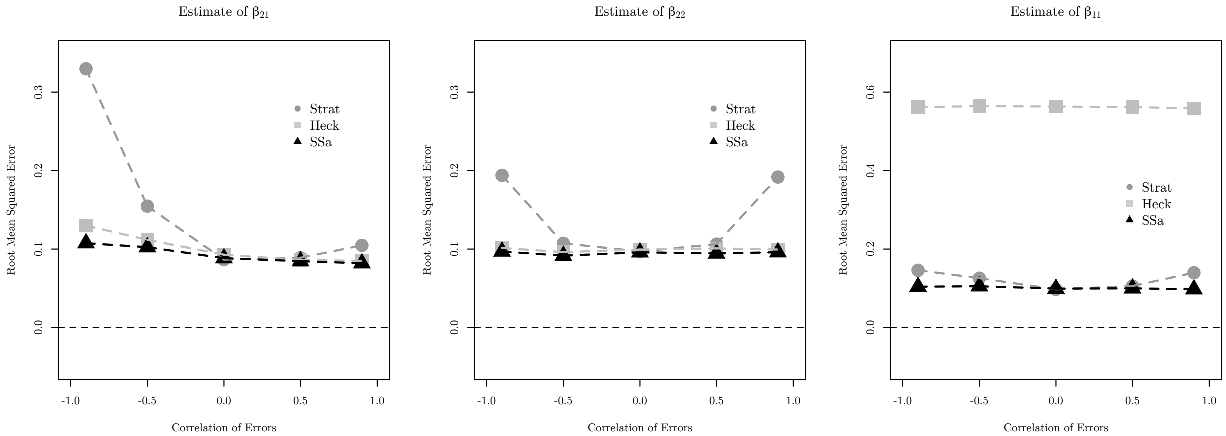
The first two theoretical expectations laid out the beginning of Figure 4.2 state that first, the strategic selection estimator (SSa) should outperform both alternative estimators (Strat, Heck) and second, that the strategic selection estimator (SSa) should outperform the strategic estimator (Strat) especially for the utility components of Player 2 (here: parameters; β_{20}, β_{21}, and β_{22}). Both of these theoretical expectations are reflected in the simulation results. The final result is that if the correlation of the two random components is 0, $\rho = 0$, the strategic estimator (Strat) as well as the strategic selection estimator (SSa) should both be unbiased as in this special case, the strategic estimator (Strat) estimator is nested in the strategic selection estimator (SSa).

Looking solely at bias may be misleading in this comparison. Since the Heck and the SSa estimators incorporate the correlation of the random components, they estimate an additional parameter. This additional parameter is costly in terms of efficiency. Hence, one might want to forego gains in unbiasedness if the efficiency costs wash those very same gains out. To see whether the above presented results hold, I show the same three plots but display this time the RMSE (root mean squared error) instead of the bias (Figure 5).

Although the RMSE is a less favorable performance measure for the SSa and the Heck we do find the same results with one exception; when the correlation of the random components in the simulation is almost 0 the standard strategic estimator (Strat) is slightly better.[17] This is an example of the bias-efficiency tradeoff as the bias in the Strat estimator is due to the small correlation (of the random components) which is itself negligible. But the efficiency gains of Strat over SSa cause a slightly better performance of the standard estimator (see left plot at correlation=0). But for all other scenarios we do find the same results as before; expect for a zero correlation the SSa estimator is the only one which is an unbiased and efficient estimator.[18]

Despite the situation in which the correlation of the error is *exactly* 0 the strategic selection esti-

Figure 5: Root Mean Squared Error (RMSE), $\frac{\sigma^2_{\epsilon_1}+\sigma^2_{\epsilon_2}}{2\cdot\sigma^2_k} = 1$



mator (SSa) is the only estimator that consistently guarantees overcoming the obstacles of strategic interaction and selection effects. This estimator is successful because it takes the selection effects into account and can uncover unbiased estimates of Player 2's parameters. Since the estimates of Player 2's parameters are unbiased, it can also outperform the strategic estimator for the coefficients of Player 1 even if the difference are smaller and more accentuated for larger correlations. Note that by picking experimental conditions in which the variance of the stochastic components are smaller or equal compared to the variance of the systematic components, the strategic estimator is at an advantage over the strategic selection estimator. If even in these cases the strategic estimator is outperformed then this should serve as a severe warning against the thoughtless use of strategic models that do not take selection into account.

4.3. *What if First Mover Is Aware of the Correlation?*

So far the results show that if the data is generated such that it is in accordance with a theoretical model based on correlated agent error the new estimator can best retrieve the estimates. The question which arises is whether the first mover is aware of this correlation or not. If the first mover is aware of the correlation, she should condition her expected probabilities of the second mover's actions on the realization of her error. Here it is now assumed that the first player is aware of the correlation and infers from his shock about the second player's shock. This is the data generating process assumed in (Signorino, 2002). The question is whether the strategic selection estimator for agent error (SSa) produces reliable estimates or not. If so, this estimator has the advantage that it is easy to use and hence offers a real alternative to explicitly accepting the possibility of selection.

The data generating process (DGP) is the same as described in Section 4.1 with the exception for Player 1's expectation of p_4. In Equation 27 p_4 is the probability of Player 2 choosing to take an action leading to outcome Y_4. When Player 1 evaluates the expected utilities from the two possible actions, Player 1 includes an expected value for p_4. To incorporate this into the DGP Player 1's shock informs that player about the second player's shock since we assume that the player knows the correlation of the two stochastic components which are unknown to the analyst. As before the

MC analysis is ran for five different levels of correlation, $\rho = -0.9, -0.5, 0, 0.5, 0.9$, and for each experimental condition 250 simulations were executed.

Figure 6: Bias and RMSE of SSa with DGP where Player 1 knows ρ

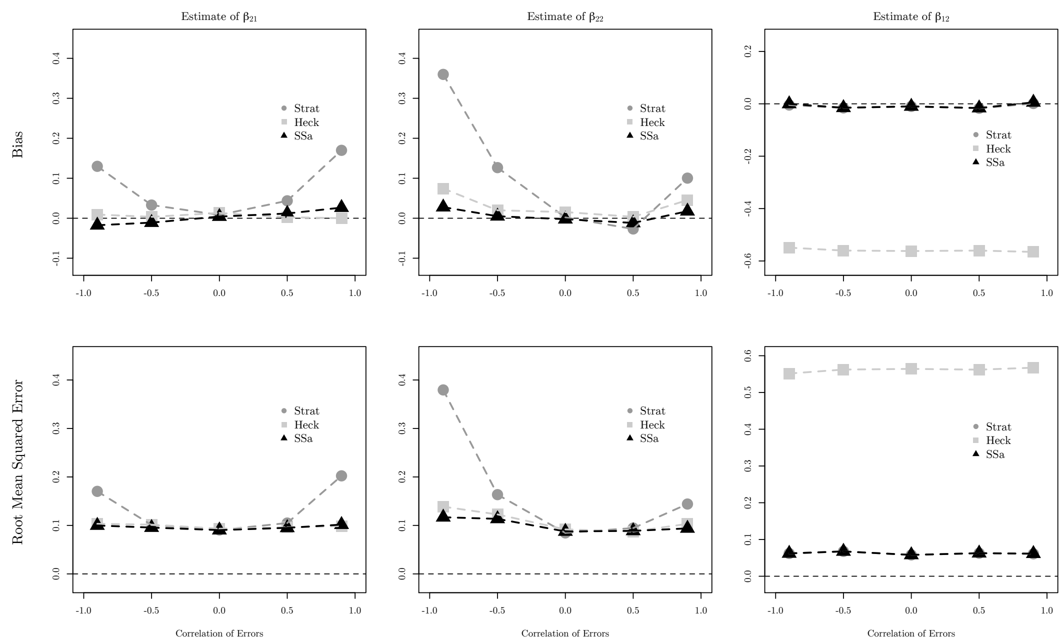


Figure 6 shows the results for the simulation and presents an almost identical picture as the results in the previous subsection. Despite the fact that the DGP is not fully identical to the one assumed to derive the statistical model the simulation results show that the estimates are almost perfect. For very high absolute values of ρ only the strategic selection estimator (SSa) is minimally biased but remains superior to both used alternatives (Strat and Heck). The implication of this result is that even when researchers assume that the first player is aware of the correlation (i.e. that the errors are correlated), the strategic selection model for agent error (SSa) allows the retrieval of good estimates. The relevance of this finding is that regardless whether one wants to make the assumption that the first player is aware of the correlated random components or not, one can still rely on this easy-to-implement estimator to obtain the estimates.

5. AN EXAMPLE OF THE STRATEGIC SELECTION ESTIMATOR

The last section showed that selection bias can be avoided by relying on the strategic selection estimator (SSa) as presented in Section 3. These theoretical results can be contrasted by an empirical example. The empirical example used here to illustrate the superiority of the strategic selection estimator (SSa) is not a rare case but rather the standard illustration of the usefulness of strategic estimators. These estimators are usually illustrated with an example modeling why states escalate conflicts and why some states capitulate and others go to war. The bulk of the data is based on the Jones et al. (1996) dataset on militarized interstate disputes (MID). The data set as used here is part of the R package Games[19] by Signorino and Kenkel (2012). I do not change the model specification but rather use the identical model as described in the vignette of the Games package. This hopefully increases the reader's confidence that this is not a rare and hand-picked example.[20]

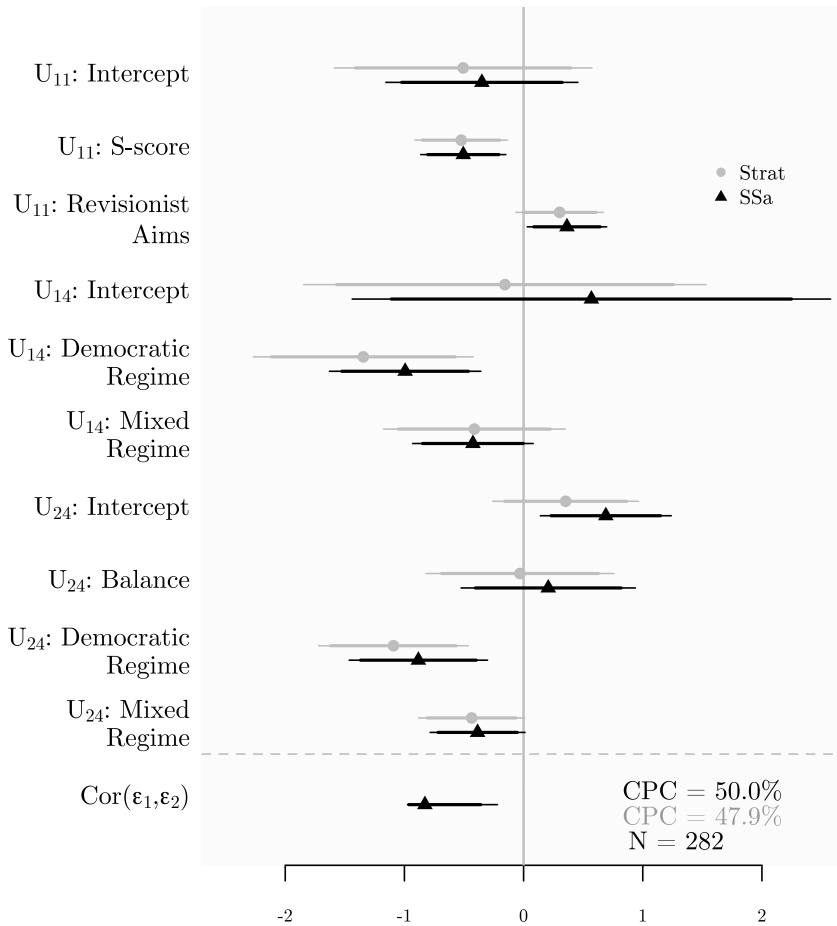
The theoretical model is as described in Figure 1 whereas state A (first player) decides whether to escalate a conflict or not. If state A escalates the conflict, state B decides whether it capitulates or goes to war. As described above, I estimate the identical model as described in Signorino and Kenkel (2012). The first player, state A, has a utility for outcome 1 (Y_1, no escalation) which is modeled through that player's S-score and whether that player has revisionist aims. The dyadic S-score is a measure of similarity between two states and is originally from Signorino and Ritter (1999). This variable measures how close states are on a political level. Hence, the larger the value, the less likely should an escalation become. Revisionist aims are measured if a country makes "claims to territory, attempting to overthrow a regime, or declaring the intention to not abide by another state's policy,... " (Jones et al., 1996: 178). The utility from outcome 3 (Y_3), escalation but state B does not chose war, is normalized to be 0. Finally, the utility for state A from escalating a conflict and having to fight a war (Y_4) is modeled by two indicators for state A's regime type which can be democratic or mixed. These indicators are based on the Polity score of any given country. The utility for player 2, which is state B, is modeled by its regime type and a balance variable which measures how strong state A is compared to state B.

The coefficient plot shows the point estimates as dots and the confidence intervals for 90% and 95%. We can start by interpreting the estimation results of an ordinary strategic model (Strat) which are shown in orange. For state A one can see that the more similar two countries are (S-score) the less likely this state is to escalate a conflict. This relationship is significant. Having revisionist aims does not have a significant impact on state A's decision to escalate. When looking at the utility for outcome Y_4, we see that escalation is less likely when state A is democratic. Turning to state B's decision one can see that the balance in military capabilities is not a significant predictor. If state B is a democracy it is significantly less likely to go to war with state A, while the effect of mixed regime is only significant for an alpha of 0.1, i.e. only the 90% confidence interval does not overlap 0.

Before looking at the estimation results of the strategic selection estimator (SSa) it is worth recapitulating the theoretical expectations. The correlation between the two systematic parts is not strong; the largest correlation between two explanatory variables is 0.15 and this suggests that bias might be found in the second actor's estimate of the intercept. We can start by looking at the estimated correlation coefficient for the two stochastic components. The point estimate is -0.83 and it is significantly different from 0.[21] The strong negative correlation causes selection bias in the estimates of the second player while none of the substantive findings change. Nevertheless, when moving up the game tree and inspecting the results of the estimates for the first player, we do find different estimates. Note also that the strategic selection estimator (SSa) fits the data better – there are 50% correctly predicted cases versus 47.9% for a model with uncorrelated errors.

Substantively, we would have made three mistakes by not taking selection bias into account. First, the direct bias would have caused us to underestimate the constant in Player 2's utility function. This means that we would have underestimated the baseline (all other covariates set to 0) probability of going to war. Second, the indirect bias would have caused us to disregard `mixed regime` (90% confidence interval) as a meaningful predictor of Player 1's behavior. Instead of uncovering that the regime type matters, we would have concluded that it only matters whether the regime is democratic or not. Thereby we would have overlooked that there is a smaller but

Figure 7: Coefficient Plot with Estimation Results



Notes: The orange estimates originate from a strategic estimator (Strat) and the purple estimates are obtained from a strategic selection estimator (SSa).

significant effect for mixed regimes versus autocracies. Finally, again by indirect bias, we would have missed the significant impact of `revisionist aims`; states which do have revisionist aims with respect to the state B are significantly more likely to escalate a conflict. This example illustrates that even if there is little correlation among the explanatory factors (which one can know *ex ante*), the biases can have an impact on the substantive claims we make. Even in this example, which is not stacked in favor of the strategic selection estimator (SSa) we find substantive differences. Hence, it is beneficial to start out with a strategic selection estimator (SSa) and only after finding an insignificant estimate for ρ resort to the pure strategic estimator (Strat).

6. CONCLUSION

This paper demonstrates that two common obstacles encountered in empirical research, the strategic nature of a DGP and selection bias, can be successfully overcome. The paper first reminds scholars that normal strategic models are sensitive to selection bias despite the fact that they explicitly model the selection stage. The paper proposes a new estimator for strategic models which does not rest on the assumption of uncorrelated errors and hence is not vulnerable to selection bias. In addition, it is shown that this estimator yields reliable results regardless of whether one assumes theoretically that the first player is aware of the correlation or not. Finally, an empirical example, which is a hard case for the proposed estimator, demonstrates the superiority of the proposed estimator.

The purpose of this paper is twofold, first, there is considerable confusion about what a strategic estimator can and cannot do (i.e. correct for selection bias) - I have shown deficiencies of this estimator when it faces data that has correlated random components. Second, I propose an easily implementable estimator that deviates from previous suggestions in the literature and liberates researchers from having to chose between the devil (selection) and the deep blue sea (strategic behavior). So far there was a theoretical options to use a strategic selection estimator proposed by Signorino (2002) which required to make strong assumptions. The problem is that its implementation is non-trivial (which might explain why it has yet not been used). This paper on the other hand proposes a different estimator which relaxes the assumptions, and is as easy to implement as ordinary strategic models.

REFERENCES

Achen, Christopher. 1986. *The Statistical Analysis of Quasi-Experiments.* Berkeley: University of California Press.

Bas, Muhammet Ali, Curtis S. Signorino, and Robert W. Walker. 2008. "Statistical Backwards Induction: A Simple Method for Estimating Recursive Strategic Models." *Political Analysis* 16(1): 21–40.

Boyes, William J., Dennis L. Hoffman, and Stuart A. Low. 1989. "An Econometric Analysis of the Bank Credit Scoring Problem." *Journal of Econometrics* 40: 3–14.

Carrubba, Clifford J., Amy Yuen, and Christopher Zorn. 2007. "In Defense of Comparative Statics: Specifying Empirical Tests of Models of Strategic Interaction." *Political Analysis* 15: 465–482.

Carson, Jamie L. 2003. "Strategic Interaction and Candidate Competition in U.S. House Elections: Empirical Applications of Probit and Strategic Probit Models." *Political Analysis* 11: 368–380.

Carson, J.L. 2005. "Strategy, selection, and candidate competition in US house and senate elections." *Journal of Politics* 67(1): 1–28.

Chen, Hsiao-Chi, James W. Friedman, and Jacques-François Thisse. 1997. "Boundedly rational Nash equilibrium: a probabilistic choice approach." *Games and Economic Behavior* 18(1): 32–54.

Clarke, Kevin A., and Curtis S. Signorino. 2010. "Discriminating Methods: Tests for Non-nested Discrete Choice Models." *Political Studies* 58: 368–388.

Fearon, James D. 1995. "Rationalist explanations for war." *International Organization* 49: 379–379.

Freedman, David A., and Jasjeet S. Sekhon. 2010. "Endogeneity in probit response models." *Political Analysis* 18(2): 138–150.

Gent, Stephen E. 2007. "Strange bedfellows: The strategic dynamics of major power military interventions." *Journal of Politics* 69(4): 1089–1102.

Greene, William H. 2003. *Econometric Analysis.* New Jersey: Pearson Education.

Hall, Peter. 2003. "Aligning Ontology and Methodology in Comparative Politics." In *Comparative Historical Analysis in the Social Sciences.* Cambridge: Cambridge University Press pp. 373–404.

Heckman, James J. 1974. "Shadow Prices, Market Wages, and Labor Supply." *Econometrica* 42: 679–694.

Heckman, James J. 1976. "The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models." *Annals of Economic and Social Measurement* 5(4): 475–492.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47(1): 153–161.

Hug, Simon, and Lucas Leemann. 2011. "Modeling the strategic interactions in representative democracies with referendums." *Presented at the Annual Meeting of the European Political Science Association, Dublin, Ireland* .

Jones, Daniel M., Stuart A. Bremer, and J. David Singer. 1996. "Militarized Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Patterns." *Conflict Management and Peace Science* 15(2): 163–213.

Kenkel, Brenton, and Curtis S. Signorino. forthcoming. "Estimating Models of Strategic Interaction in R." *Journal of Statistical Software* .

Leblang, David. 2003. "To devalue or to defend? The political economy of exchange rate policy." *International Studies Quarterly* 47(4): 533–560.

Lewis, Jeffrey B., and Kenneth A. Schultz. 2003. "Revealing Preferences: Empirical Estimation of a Crisis Bargaining Game with Incomplete Information." *Political Analysis* 11: 345–367.

Maddalla, G.S. 1983. "Limited-Dependent and Qualitative Variables in Econometrics." *Cambridge: Cambridge University Press* 999: econometrics.

Manski, Charles F. 1977. "The Structure of random Utility Models." *Theory and Decision* 8: 229–254.

McKelvey, Richard D., and Thomas R. Palfrey. 1995. "Quantal Response Equilibria for Normal Form Games." *Games and Economic Behavior* 10: 6–38.

McKelvey, Richard D., and Thomas R. Palfrey. 1996. "A Statistical Theory of Equilibrium in Games." *The Japanese Economic Review* 47: 186–209.

McKelvey, Richard D., and Thomas R. Palfrey. 1998. "Quantal Response Equilibria for Extensive Form Games." *Experimental Economics* 1: 9–41.

Reed, William. 2000. "A Unified Statistical Model of Conflict Onset and Escalation." *American Journal of Political Science* 44(1): 84–93.

Sartori, Anne E. 2003. "An Estimator for Some Binary-Outcome Selection Models Without Exclusion Restrictions." *Political Analysis* 11: 111–138.

Signorino, Curtis S. 1999. "Strategic Interaction and the Statistical Analysis of International Conflict." *American Political Science Review* 93: 279–297.

Signorino, Curtis S. 2002. "Strategy and Selection in International Relations." *International Interactions* 28(January - March): 93–115.

Signorino, Curtis S. 2003*a*. *STRAT – A Program for Analyzing Statistical Strategic Models Version 1.4*. Department of Political Science, University of Rochester.

Signorino, Curtis S. 2003*b*. "Structure and Uncertainty in Discrete Choice Models." *Political Analysis* 11: 316–344.

Signorino, Curtis S., and Ahmer Tarar. 2006. "A unified theory and test of extended immediate deterrence." *American Journal of Political Science* 50(3): 586–605.

Signorino, Curtis S., and Brenton Kenkel. 2012. *Package 'games'*. R: A language and environment for statistical computing.

Signorino, Curtis S., and Jeffrey M. Ritter. 1999. "Tau-b or not tau-b: Measuring the similarity of foreign policy positions." *International Studies Quarterly* 43: 115–44.

Signorino, Curtis S., and Kuzey Yilmaz. 2003. "Strategic Misspecification in Regression Models." *American Journal of Political Science* 47(3): 551 – 566.

Smith, Alastair. 1999. "Testing Theories of Strategic Choice: The Example of Crisis Escalation." *American Journal of Political Science* 43(4): 1254–1283.

Vuong, Quang H. 1989. "Likelihood ratio tests for model selection and non-nested hypotheses." *Econometrica* 57(2): 307–333.

Notes

[1]See Reed (2000) for a non-strategic proposition focussing on selection effects for international relations. See Smith (1999) for an alternative argument regarding strategy and selection.

[2]There are three common assumptions pertaining to the origin of the uncertainty: agent error, private information about payoffs, and regressor error.

[3]This is identical to the random utility model derivation for more prevalent models as the probit model. For the probit model one also ends up making a distributional assumption ($\varepsilon \sim N(0,1)$) for the difference of two random variables.

[4]See e.g. equation 2 in Kenkel and Signorino (forthcoming).

[5]A further assumption is the constant variance of estimators. But this might not hold as e.g. moments of crisis could both affect the first and second actor and increase the variance of their errors. But, Signorino and Kenkel (2012) provide an estimator that allows the analyst to parametrize the variance term (similar to heteroscedastic probit) and this can mitigate the problem as long as one can condition on observable covariates.

[6]I am grateful to an anonymous reviewer for pushing me on this point.

[7]The reason for this can be seen when one derives the expected value of y_i conditional on observing y_i: $E(y_i|z_i^*) = E(y_i|y_i \text{ is observed})$ and $E(y_i|z_i^*) = \boldsymbol{\beta}'\mathbf{x}_i + \rho\sigma_\epsilon \left(\phi\left(\frac{\boldsymbol{\gamma}'\mathbf{w}_i}{\sigma_u}\right)/\Phi\left(\frac{\boldsymbol{\gamma}'\mathbf{w}_i}{\sigma_u}\right)\right)$ where ρ is the correlation between the two random parts ($\rho = \frac{\sigma_{\epsilon,u}}{\sigma_\epsilon\sigma_u}$). See chapter 9 of Maddalla (1983) for an in-depth treatment of selection.

[8]The table in the appendix lists those examples.

[9]This detail can be found when looking at the actual code. I am thankful to Curt Signorino for sharing his code with me.

[10]See e.g. Carrubba, Yuen, and Zorn (2007, 474): "*In his 'International Interactions' paper, Signorino (2002) performs a simulation of a strategic model with correlated errors but concludes that such a model is extremely complicated and recommends using an estimator that correctly captures the strategic behavior at the sacrifice of modeling the correlated errors. Thus, because Signorino never promotes estimating a model with strategic behavior and correlated errors, in this study we focus on the model with uncorrelated errors.*" (Carrubba, Yuen, and Zorn, 2007, 474) or also Carson (2003, 374): *"Indeed, failing to model strategic interaction between two or more actors can lead to a form of model misspecification far worse than failing to account for selection bias that results from correlated disturbances (Signorino 2002)."*.

[11]Note, that in (Signorino, 2002) there is no comparison (in terms of RMSE, bias, or alike) between the strategic estimator and the strategic selection estimator. The only comparison presented is between the non-strategic Heckman selection model and the strategic estimator with uncorrelated errors (Strat).

[12]Unlike several different versions of the strategic estimator, there is no canned routine or simple procedure for the strategic selection estimator that Signorino (2002) presented. Neither in the original software `STRAT` (Signorino, 2003*a*) nor in subsequent `R` package `games` (Signorino and Kenkel, 2012) is an implementation of the strategic selection estimator (last checked on 1/27/2012). Due to the correlated structure of its random components one cannot rely on statistical backwards induction (Bas, Signorino, and Walker, 2008).

[13]The MC results are presented for both potential DGPs, where the first actor knows that the random components are correlated and also for the case where the first actor does not know that they are correlated.

[14]All applications of a strategic estimator (Strat) that I have found contain two sets of variables which do not fully overlap.

[15]There are two possibilities to estimate this model, one can use full information maximum likelihood (FIML) or statistical backwards induction (SBI) (Bas, Signorino, and Walker, 2008). Here, I rely on FIML as it allows me to estimate the parameters and standard errors in one step.

[16]Here I present the results for equal ratio. The appendix (Section 7) presents the remaining figures for the ratio being 0.1, 0.5, and 2.0.

[17]The correlation of the errors is never exactly 0 in finite samples. The difference between Strat and SSa in the left panel is hard to see, RMSE(Strat)$= 0.0866$ and RMSE(SSa)$= 0.0884$.

[18]In the appendix I show the distribution of the estimated correlation coefficients (Figure 11). That plot shows that the estimated correlations are close to the true value. if the true correlation is 0 the estimates are between -0.110 and 0.099. Hence, there is little danger of claiming a substantive selection bias when in fact the true value of the correlation is 0.

[19]`http://cran.r-project.org/web/packages/games/games.pdf`, see also Kenkel and Signorino (2012).

[20]One of the consequences is that unlike the full MID data set, this data set only covers the years 1800-1899.

[21]The confidence intervals are asymmetric because the uncertainty is determined with simulation (instead of the Delta method) and there is truncation due to the non-linear function of the reparametrization of ρ.

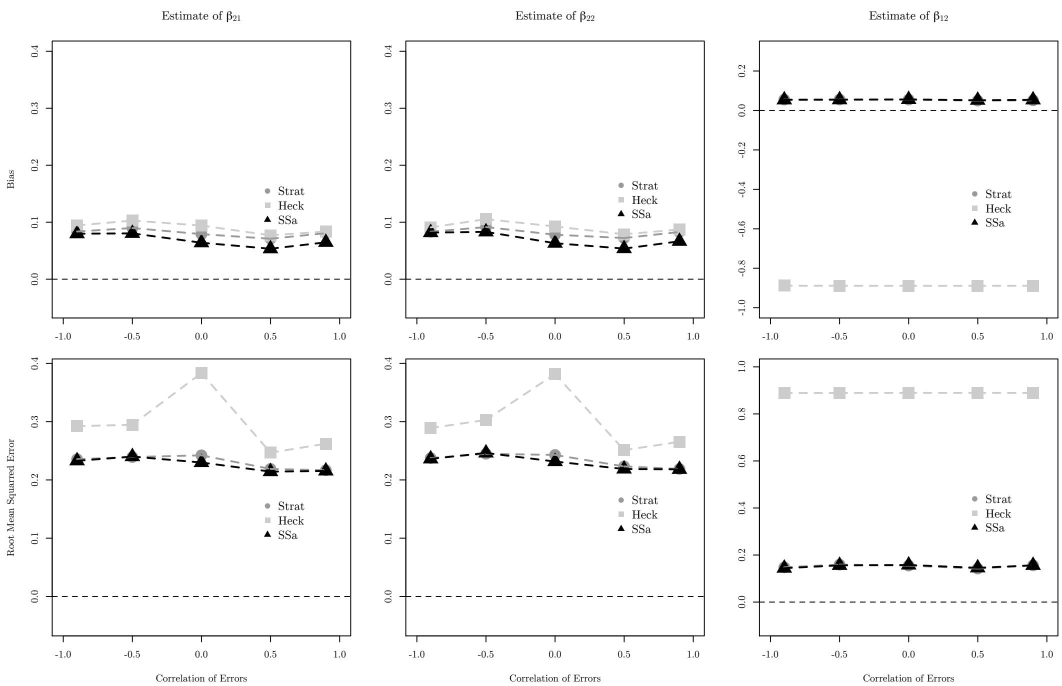
7. APPENDIX FOR READERS

7.1. *Additional Results from the MC Analysis*

In subsection 4.2 I present a number of plots based on the Monte Carlo analysis. The specific results which are presented in Figure 3, Figure 4, and Figure 5 are all based on an experimental set-up in which the sum of the two error variance is equal to the sum of the variances of the systematic parts. As mentioned in subsection 4.1 I also varied this condition such that $(\sigma^2_{\epsilon_1} + \sigma^2_{\epsilon_2})/(2 \cdot \sigma^2_k)$ takes on the values $0.1, 0.5$, and 2.0.

In Figure 8, Figure 9 and Figure 10 we see the bias and RMSE (root mean squared error) for each of the three estimators under different degrees of error correlation. Both figures represent the results for experimental conditions where the error variance is much smaller than the variance of the systematic parts. Theoretically, when the error variance goes to zero the advantage of using a strategic selection estimator (here: SSa) goes to zero, because the selection bias goes to zero.[22]

Figure 8: Bias and RMSE for $\frac{\sigma^2_{\epsilon_1}+\sigma^2_{\epsilon_2}}{2\cdot\sigma^2_k} = 0.1$

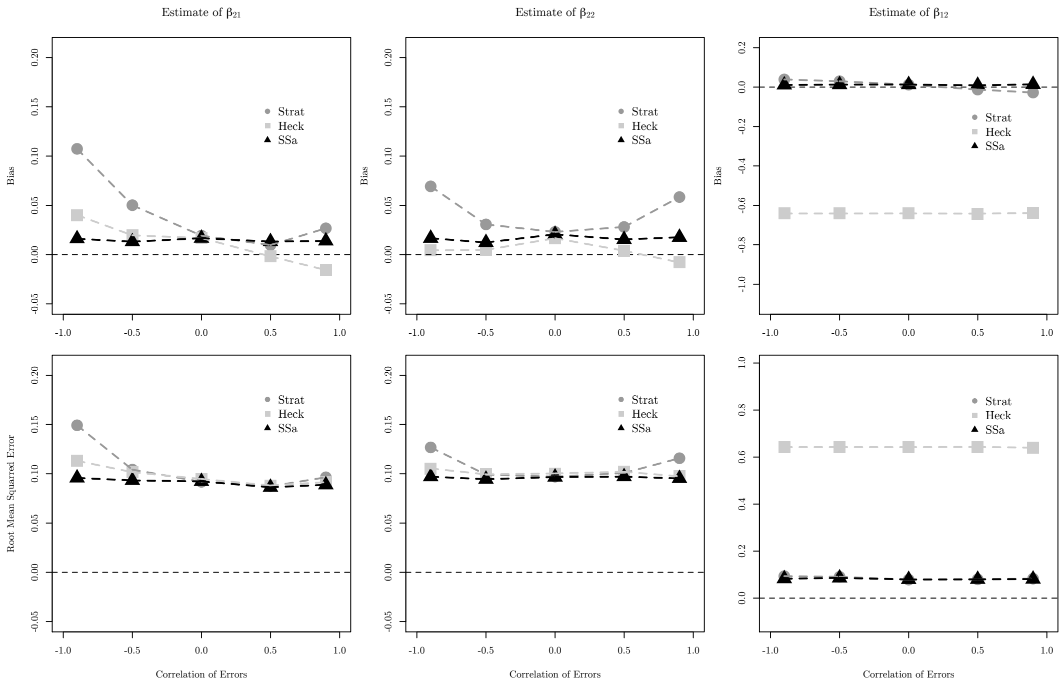


First, Figure 8, Figure 9 and Figure 10 show essentially the same results as Figure 3, Figure 4, and Figure 5 did. The strategic selection estimator (SSa) is at least as good as the strategic estimator (Strat). When the variance of the stochastic component becomes very small ($\frac{1}{10}$'th of the variance of the systematic part) the strategic selection estimator (SSa) is on par with the strategic estimator, since the selection bias becomes negligible. As soon as the error variance increases relative to the variance of the systematic part, the strategic selection estimator (SSa) outperforms the other two estimators (Heck, Strat).

Once variance of the stochastic components increases to half of the variance of the systematic components the resulting picture is as expected. While the Heckman selection estimator (Heck) delivers sufficiently good estimates for the second player's coefficients it is way off for the first player's coefficient. Similarly, the standard strategic estimator (Strat) does not perform well for the second player's coefficients. Again, the only estimator which guarantees under all circumstances the lowest RMSE is the strategic selection estimator (SSa).

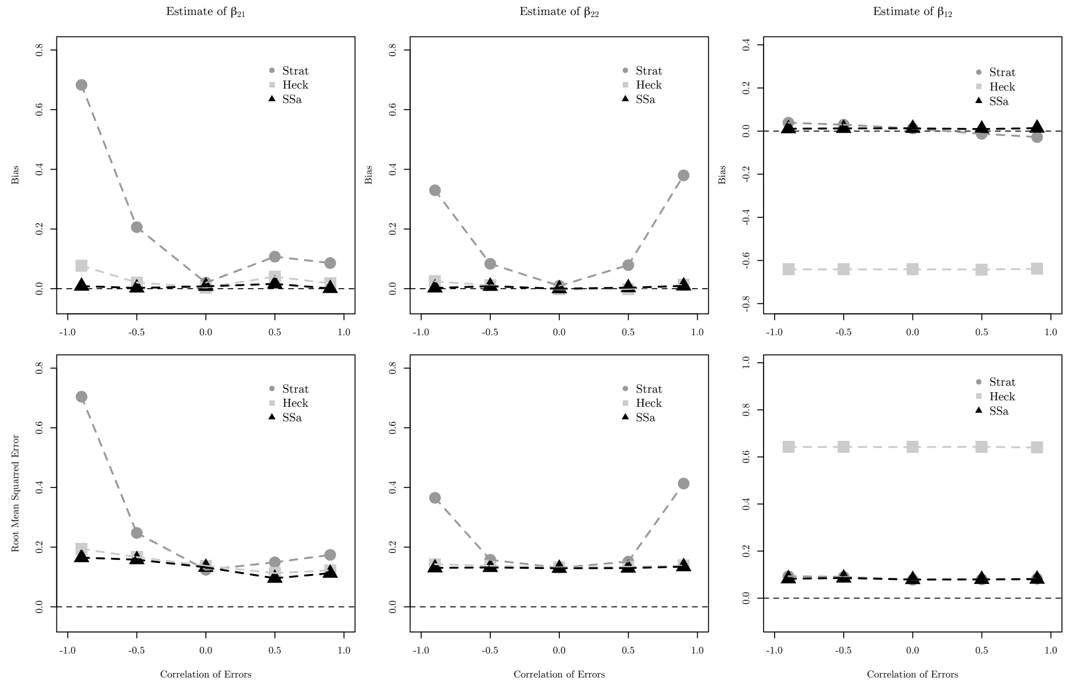
Finally, once the error variance is larger than the variance of the systematic part the results are even

Figure 9: Bias and RMSE for $\frac{\sigma^2_{\epsilon_1}+\sigma^2_{\epsilon_2}}{2\cdot\sigma^2_k} = 0.5$



clearer than before. The bias in the strategic estimates for the second player's coefficient is even greater than before and the Heckman model (Heck) keeps delivering almost unbiased results. But for the first player's coefficient this is reversed. As with all other experimental settings, the only estimator which delivers consistently the lowest RMSE and lowest bias is the strategic selection estimator (SSa).

Figure 10: Bias and RMSE for $\frac{\sigma^2_{\epsilon_1}+\sigma^2_{\epsilon_2}}{2\cdot\sigma^2_k} = 2.0$



7.2. *Estimation of the Correlation*

The following histograms show the distribution of the estimates for the correlation coefficient ($\hat{\rho}$). The first row is for a small sample of 500 observations, the middle row for 1,000 observations, and the bottom row shows the results for a large sample of 2,000 observations. From this plot we gain the insight that the estimates of the correlation coefficient are very precise and that there is little to no risk of mistakenly estimating a large value for ρ when in fact this ρ is 0. Evaluating bivariate normals can pose problems as shown previously by (Freedman and Sekhon, 2010). Despite this, the estimator has performed well in the simulations and the estimates of the correlation have not been nonsensical. Very small values of $\hat{\rho}$ can be due to a true $\rho = 0$ but in this case the strategic estimator (Strat) as well as the strategic selection estimator (SSa) produce almost identical results.

Figure 11: Histogram of the Estimates for ρ