New Results on the Identification of Stochastic Bargaining Models*

Antonio Merlo† Xun Tang‡

February 15, 2017

Abstract

We present new identification results for stochastic sequential bargaining models when data only reports the time of agreement and the evolution of observable states. With no information on the stochastic surplus available for allocation or how it is allocated under agreement, we recover the latent surplus process, the distribution of unobservable states, and the equilibrium outcome in counterfactual contexts. The method we propose, which is constructive and original, can also be adapted to establish identification in general optimal stopping models.

Key words: Nonparametric identification, stochastic sequential bargaining
JEL codes: C14, C73, C78

*We thank Michael Shashoua for capable research assistance. This research is funded by NSF Grant #SES-1448257.

†Email: amerlo@rice.edu. Economics Department, Rice University MS-22, Houston, TX 77251-1892.

‡Email: xun.tang@rice.edu. Economics Department, Rice University MS-22, Houston, TX 77251-1892.

1 Introduction

The model of stochastic sequential bargaining (SSB) proposed by Merlo and Wilson (1995, 1998) has been widely used to study negotiations in a broad range of environments. In this model, the surplus to be allocated among players (or the "cake") and the order in which players make offers and counteroffers (or the "bargaining protocol") evolve over time according to a stochastic process. Within a framework of bargaining with complete information, the model rationalizes delays in reaching agreement, which are often observed in actual negotiations. Also, the model has a unique stationary subgame perfect equilibrium, which is stochastic and can be computed as the solution of a fixed-point problem.

For these reasons, the SSB framework naturally lends itself to estimation and has been used in a variety of empirical applications that range from the formation of coalition governments in parliamentary democracy (Merlo (1997); Diermeier, Eraslan and Merlo (2003)), to collective bargaining agreements (Diaz-Moreno and Galdon (2004)), to corporate bankruptcy reorganizations (Eraslan (2008)), to the setting of industry standards in product markets (Simcoe (2012)), and to sovereign debt renegotiations (Benjamin and Wright (2009); Bi (2008); Ghosal, Miller and Thampanishvong (2010)).

In Merlo and Tang (2012), we study the identification of a general class of SSB models under minimal parametric restrictions on their structural elements given different data scenarios. For the "complete data" scenario (where econometricians observe the size of the cake in each period regardless of whether an agreement is reached), and the "incomplete data with censored cakes" scenario (where econometricians only observe the size of the cake in the period when an agreement is reached), Merlo and Tang (2012) provide positive results for identifying the structural elements of these models. However, for the "incomplete data with unobservable cakes" scenario (where econometricians only observe the timing of agreement, but never observe the size of the cake), Merlo and Tang (2012) only show partial identification of counterfactual outcomes (i.e., the probability of reaching an agreement conditional on observed states), but do not establish point-identification of model elements.

In this paper, we address the issue of nonparametric point-identification of SSB models under the incomplete data scenario with unobservable cakes. Empirical contexts of stochastic bargaining models typically differ in what econometricians observe in the data, and this particular scenario has great empirical significance. There are many examples of bargaining situations where the observed data only contains information about the evolution of states and the duration of negotiation, but not the size or the agreed upon allocation of the surplus. For example, Simcoe (2012) uses a SSB model to analyze the adoption of standards by the Internet Engineering Task Force. He uses an administrative data set that only reports the time between initial submission and final revision of 2,601 Internet Protocols between 1993 and 2003. Similarly, Diaz-Moreno and Galdon (2004) use a SSB model to study collective bargaining agreements in Spain. The data set they use only contains the duration of 545 collective wage negotiations in Spain in the late 1980s. Eraslan (2008) uses a SSB model to estimate the liquidation value of 128 firms which filed for Chapter 11 bankruptcy in the United States between 1990 and 1997. In this application the claimants of firms' assets (debt- and equity-holders) bargain over the allocation of a firm's reorganization value after it emerges from bankruptcy. Eraslan (2008) compiles a U.S. Corporate Bankruptcy data set which reports the duration as well as the evolution of relevant states in all negotiation processes.

We exploit the identifying power from an additive state to recover the latent surplus process, the distribution of unobserved states, and the outcome in counterfactual experiments. Broadly speaking, an additive state is defined as a state variable whose marginal effect on the latent total surplus is known (up to a scale normalization). Our method requires three identifying conditions on the additive state: (a) its Markovian transition is separable from that of the other observed states (this holds, for example, if it evolves independently from the other states); (b) its transition to the next period demonstrates a pattern with limited feedback; and (c) it is conditionally independent from the unobserved state, or the error term. Each of these conditions is empirically plausible, and holds in a variety of specifications commonly used in the empirical literature. The conditions (a) and (b) were not assumed in Merlo and Tang (2012), and are instrumental for our new identification results in this paper.

Our approach for identification exploits the fact that in the unique stationary subgame perfect equilibrium of a SSB model, an agreement is reached only in a period when the realized total surplus exceeds the discounted continuation value (or the conditional expectation of total equilibrium payoff in the next period given current states). We establish the identification results through sequential steps. First, using conditions (a) and (b) above, we identify the marginal effect of the additive state on the continuation value in equilibrium. Next, under condition (c) we use the variation in the additive state and knowledge of its marginal effect on the continuation value to recover the error distribution. Then, we substitute the elements identified from these two steps into the Bellman equation which characterizes the equilibrium continuation value as a fixed point. This leads to a "quasi-structural" form of the Bellman equation, which allows us to recover the latent surplus process up to a location parameter in the error distribution. We then propose two parallel alternatives, one of which relies on the large support of the additive state while the other on the sufficient exogenous variation in the transition of states, to jointly identify the location parameter and the surplus process. Last, we show how to extend our method to make valid inference of the counterfactual probability for an agreement even when these two identifying conditions are relaxed.

Our method can be adjusted to identify a broad class of nonparametric optimal stopping models where neither the error distribution nor the per-period utility is parametrically specified. We provide the details in Appendix B. The identifying argument works for optimal stopping models because an agreement in a stationary subgame perfect Nash equilibrium in the SSB model satisfies a "separation principle" (i.e., it depends only on the current states but not the identity of the proposer). Thus the characterization of an agreement in equilibrium is analogous to that of individual rationality in a single-agent optimal stopping problem.

Our method is partly motivated by the concept of a "special regressor" in static qualitative response models. In particular, our additive states inherit two key properties of a special regressor: additive separability in the latent utility, and conditional independence from unobserved errors. Lewbel (2000) shows how to use special regressors with a large support to recover the latent utility in static single-agent binary response models nonparametrically. Chen, Khan and Tang (2015) quantify the identifying power of the special regressor in static single-agent binary response models under different location normalizations using two criteria: Fisher information and the proportion of regressor supports that is used for identification. Lewbel and Tang (2015) extend the special regressor argument to identify a static simultaneous binary game with incomplete information. Berry and Haile (2014) use related exclusion restrictions to identify a market demand system for differentiated products.

In comparison with these papers, our contribution is to provide positive identification results in a substantially different model of a dynamic game. In this dynamic environment, the additive states evolve over time and affect the equilibrium outcome through its non-additive, non-linear impact on the optimal continuation value. The latter is characterized as a solution of a complex fixed-point problem, thus complicating the identification question. The solution we propose in this paper taps into some additional structure that is not explored previously in the existing literature: separability and limited feedback in the transition of additive states.

Section 2 introduces the canonical model of stochastic sequential bargaining and defines its equilibrium. Section 3 presents the identification method via several steps. Section 4 discusses identification and counterfactual analyses when some key assumptions are relaxed. Section 5 concludes. Proofs not presented in the text are included in Appendix A. In Appendix B we provide details about how to adjust our method to identify a nonparametric optimal stopping model.

2 The Model

Consider an infinite-horizon bargaining game with $N \geq 2$ players (indexed by $i = 1, ..., N$) who share the same constant discount factor $\beta \in (0, 1)$. In each period $t = 0, 1, ...,$ all players observe a vector of states (X_t, ϵ_t) with support $\mathcal{X} \times \mathcal{E}$. Throughout the paper, we use calligraphic letters $\mathcal{R}$ to denote the marginal support of a generic random vector R; we denote the history of a random vector up to period t by $R^t \equiv \{R_0, R_1, ., R_t\}$. For a generic pair of random vectors R_1, R_2, let $F_{R_2|R_1}$ denote the distribution of R_2 conditional on R_1, and write it as $F(r_2|r_1)$ while evaluating it at realized values $R_1 = r_1, R_2 = r_2$.

In each period t, a vector of states (X_t, ϵ_t) determine the set of feasible payoff vectors to be allocated, which is denoted by $\mathcal{C}(X_t, \epsilon_t) \equiv \{u \in \mathbb{R}_+^N : \sum_{i=1}^N u_i \leq c(X_t, \epsilon_t)\}$, with $c : \mathcal{X} \times \mathcal{E} \to \mathbb{R}_+$ determining the total surplus, or the "cake", to be shared.[1] In each period, a player i is randomly selected to be the proposer with probability $L_i \in (0, 1)$ such that $\sum_{i=1}^N L_i = 1$. Let $L \equiv \{L_1, ..., L_N\}$. We denote the random identity of the proposer in period t by κ_t and denote its realization by $\Bbbk_t$. The data available to researchers report X_t but not the scalar noise ϵ_t. We maintain the following restriction on the transition of states throughout the paper.

Assumption 1 *The transition of states over time is such that*

$$\Pr\{X_{t+1} \leq x_{t+1}, \epsilon_{t+1} \leq \varepsilon_{t+1} | X^t = x^t, \epsilon^t = \varepsilon^t) = F(\varepsilon_{t+1}|x_{t+1})G(x_{t+1}|x_t) \text{ for all } t \quad (1)$$

where the error distribution F and the transition of observable states G are both time-homogeneous.

We use R and R' to denote random vectors in the current and the next period respectively. Assumption *1* requires that the dynamics between current and subsequent states (X, ϵ) and (X', ϵ') be captured by the transition of states observed in the data. This is a condition that has been widely used in industrial organization and labor economics (e.g. Rust (1987) and Rust (1994)).

At the beginning of each period, players observe a vector of realized states (x, ε) and are informed of the realized identity of the proposer $\Bbbk$. The proposer then chooses to either suggest an allocation in $\mathcal{C}(x, \varepsilon)$ or to pass and let the game move to the next period. If he proposes an allocation, the other players respond by either accepting or rejecting the proposal. If no proposal is offered or the proposal is rejected by some player, the game moves to the next period where new states (x', ε') are realized and a new proposer $\Bbbk'$ is selected with probabilities L. The procedure is then repeated with total surplus given by $c(x', \varepsilon')$. This game continues until an allocation is proposed and unanimously accepted (if ever).

The parameters $\{\beta, c, L, G, F_{\epsilon|X}\}$ are common knowledge among all players. Let $S_t \equiv (X_t, \epsilon_t, \kappa_t)$ denote the information available to the players at time t. Given any initial realized

[1] This environment assumes that the players have time-separable quasi-linear von Neumann-Morgenstern utility functions over the commodity space, and that a good with constant marginal utility to each player (e.g., money) can be freely transferred. In the terminology of Merlo and Wilson (1995, 1998), this environment is defined as a stochastic bargaining model with transferable utility.

information S_0, an *outcome* of the bargaining game (τ, η_τ) consists of a stopping time τ and a random N-vector η_τ (measurable with respect to S^τ) such that $\eta_\tau \in \mathcal{C}(X_\tau, \epsilon_\tau)$ is a feasible division of surplus under the state (X_τ, ϵ_τ) if $\tau < +\infty$, and $\eta_\tau = 0$ if $\tau = +\infty$. Given a realized sequence of information $(s_0, s_1, ..)$, τ is the period in which a proposal is accepted by all players, and η_τ is the accepted proposal when the state is (X_τ, ϵ_τ) and the identity of the proposer is κ_τ. For a game starting with initial states (x, ε) and a proposer $\Bbbk$, an outcome (τ, η_τ) implies a von Neumann-Morgenstern payoff to player i, $E[\beta^\tau \eta_{\tau,i} | S_0 = (x, \varepsilon, \Bbbk)]$, where $\eta_{\tau,i}$ is the i-th component in η_τ.

A *stationary outcome* is such that there exists a measurable subset $\mathcal{S}^\mu$ of the support of S, and a measurable function $\mu : \mathcal{S}^\mu \to \mathbb{R}^N_+$ such that (i) $S_t \notin \mathcal{S}^\mu$ for all $t = 0, 1, ., \tau - 1$; (ii) $S_\tau \in \mathcal{S}^\mu$; and (iii) $\eta_\tau = \mu(S_\tau)$. That is, no allocation is implemented until some state and proposer identity $s = (x, \varepsilon, \Bbbk) \in \mathcal{S}^\mu$ is realized, in which case a feasible proposal $\mu(s)$ is accepted by all players. Given property (iii), we let $v^\mu(s) \equiv E[\beta^\tau \mu(S_\tau) | S_0 = s]$ denote the vector of individual von Neumann-Morgenstern payoffs given the initial state and the proposer identity in s. It follows from the definition of stationary outcome that $v^\mu(s) = \mu(s)$ for all $s \in \mathcal{S}^\mu$ and $v^\mu(s) = E[\beta^\tau \mu(S_\tau) | S_0 = s]$ for all $s \notin \mathcal{S}^\mu$. Hence, a stationary outcome is characterized by the triple $\{\mathcal{S}^\mu, \mu, \tau\}$.

A history up to time t is a finite sequence of realized states, identities of proposers, and actions taken up to time t. A strategy for player i specifies a feasible action at every history at which he must act. A strategy profile is a measurable N-tuple of strategies, one for each player. At any history, a strategy profile induces an outcome and hence a payoff for each player. A strategy profile is a *subgame perfect equilibrium* (SPE) if, at every history, it is a best response to itself. We refer to the outcome and payoffs induced by a subgame perfect strategy profile as an SPE outcome and SPE payoff respectively. A strategy profile is *stationary* if the actions prescribed at any history depend only on the current state, the proposer identity, and the offer. A stationary SPE (SSPE) outcome and payoff are the outcome and payoff generated by a subgame perfect strategy profile which is stationary.

The following lemma characterizes the players' actions and outcomes in the unique SSPE of the game. Let $v_i : \mathcal{S} \to \mathbb{R}_+$ denote the SSPE payoff for player i, $v = (v_1, ..., v_N)$ the SSPE payoff vector, and $w = \sum_{i=1}^N v_i$ the SSPE total payoff of all players in the bargaining game. Let $\mathcal{F}^N$ denote the set of bounded measurable functions mapping from $\mathcal{S}$ to $\mathbb{R}^N_+$.

Lemma 1 *(Lemma 1 in Merlo and Tang (2012).) Suppose Assumption 1 holds. Then (a) $v \in \mathcal{F}^N$ is the unique SSPE payoff vector if and only if $v = A(v)$, where for any s and $i = 1, ..., N$, the operator $A(.)$ is defined as:*

$$A_i(v(s)) \equiv \max\left\{c(x, \varepsilon) - \beta E\left[\left.\textstyle\sum_{j \neq i} v_j(S')\right| x\right], \beta E\left[\left. v_i(S')\right| x\right]\right\} \text{ for } i = \Bbbk, \tag{2}$$

$$A_i(v(s)) \equiv \beta E[v_i(S') | x] \text{ for } i \neq \Bbbk; \tag{3}$$

(b) the SSPE total payoff w does not depend on the identity of the proposer, and solves

$$w(x, \varepsilon) = \max\{c(x, \varepsilon), \beta E[w(X', \epsilon')|x]\}; \tag{4}$$

and (c) an agreement is reached under the state (x, ε) if and only if $c(x, \varepsilon) \geq \beta E[w(X', \epsilon')|x]$.

The total payoff and the occurrence of agreement in SSPE only depend on the current states (X, ϵ), but not the proposer identity κ. This property, also known as the "separation principle", is key to our identification results. In contrast, the individual SSPE payoffs $\{v_i\}_{i=1}^{N}$ do depend on the identity of the proposer. Only $i = \Bbbk$ claims the additional "gain to the proposer" $c(x, \varepsilon) - \beta \sum_j E[v_j(S')|x]$ in addition to her continuation payoff $\beta E[v_i(S')|x]$; all others just get their continuation payoffs. Under Assumption 1, the expectation of next period's SSPE total payoff is independent of ε conditional on x.

If an agreement is reached in any period, the proposer extracts any surplus over what the players obtain by delaying agreement until the next period (i.e. their equilibrium continuation payoffs). The separation principle implies that the identity of the proposer affects how the cake is allocated, but not the states in which it is allocated. Furthermore, the gains from proposing are also independent of the identity of the proposer, and the characterization of the SSPE total payoff is equivalent to the solution of the single-agent optimal stopping problem of deciding when to consume a stochastic cake. The SSPE total payoff maximizes the expected discounted size of the cake (i.e. the discounted expectation of total surplus allocated among the players). The unique SSPE of the game, and hence any delay in agreement, is therefore Pareto efficient.

The fact that a temporary delay in agreement is a possible equilibrium outcome follows from the possibility that the discounted size of the cake need not decline in every period. In particular, equilibrium delays occur in states where the cake is "too small": that is, the sum of the continuation payoffs of all players (including the proposer) exceeds the current cake.

3 Identification

We are interested in recovering the model structure $\{c, F_{\epsilon|X}, G\}$, using the distribution of states and the probability for a unanimous agreement from a data that contains a large sample of bargaining games. For each bargaining game in the data, we assume that researchers observe the time to agreement and the history of the states X, but not ϵ. The discount factor β is assumed known to researchers. Also, we posit that all the bargaining games observed in the data share the same state transition probabilities and the same cake function $c : \mathcal{X} \times \mathcal{E} \to \mathbb{R}_+$, and all players follow SSPE strategies. On the other hand, the number of players is allowed to vary across bargaining games, and may or may not be reported in the data.

To begin, we summarize the structural link between model elements and the states and outcomes reported in the data. Let $\pi(x) \equiv E[w(X', \epsilon')|x]$ denote the continuation value (the ex ante total payoff in SSPE), which does not depend on ε due to Assumption 1. Lemma 1 implies that in equilibrium an agreement occurs under state x if and only if

$$c(x, \varepsilon) \geq \beta\pi(x), \tag{5}$$

where $\pi(.)$ solves

$$\pi(x) = \int \left[\beta\pi(x') + \int \max\{c(x', \varepsilon') - \beta\pi(x'), 0\} dF(\varepsilon'|x')\right] dG(x'|x) \tag{6}$$

for all $x \in \mathcal{X}$.

It is worth emphasizing that we consider a scenario where the data reports the evolution of observed states and whether an agreement is reached in each period, or a round of negotiation, but does *not* report the size of total surplus to be shared *or* the accepted allocations. Nor does the data report the identity of the proposer. The goal of this section is to show how the model elements can be identified in this scenario under various conditions.

Our identification method uses the probability for an agreement conditional on the observed states in a SSPE. The observation of the duration of negotiation per se is not essential for our method. In some environments, a round of negotiation may be directly defined in the data; in others, it can be proxied by the frequency of observations, or assumed using some institutional facts. Also, our method does not require knowledge of the distribution of the proposer identity L_i, even though it can be recovered if the proposer identity is reported in each round of negotiation.

3.1 Additive state and its marginal effect on continuation value

Our identification strategy exploits the variation in an additive state variable whose marginal effect on the latent surplus is known to the researchers.[2] We formalize the restrictions on such a state as follows.

Assumption 2 *The vector x is partitioned into $x = (x_1, x_2)$ so that (i) x_2 is a scalar variable and $c(x, \varepsilon) = c_0(x_1) + x_2 + \varepsilon$ for all x; (ii) for all x, $F(\varepsilon|x) = F(\varepsilon|x_1)$ and $F(0|x_1) = 0.5$; (iii) $G(x'|x) = G_1(x'_1|x_1)G_2(x'_2|x_2)$; and (iv) for any generic function h that is continuous and bounded over $\mathcal{X}$, both $\int \max\{c(x, \varepsilon) - h(x), 0\} dF(\varepsilon|x)$ and $\int h(x')dG(x'|x)$ are continuous and bounded over $\mathcal{X}$.*

[2]The term "additive state" is potentially misleading. In fact, our identification strategy is valid under the more general case where the partial effect of x_2 on c is non-constant but is known to the researcher over $\mathcal{X}$. Nevertheless, we focus on an "additive state" throughout the paper only for the sake of simple exposition.

Part (i) in Assumption 2 is less restrictive than it appears. It is common for empirical papers to adopt an additive error and a linear specification of the median surplus function c, in which case that the marginal effect of the additive X_2 equals 1 is just a scale normalization (because we are not assuming any parametric form or scale for the error distribution $F_{\epsilon|X}$).

Part (ii) combines a conditional independence condition ($X_2 \perp \epsilon \mid X_1$) with a median independence condition on the error term; it also implements an innocuous location normalization of zero median. Under (i) and (ii), the probability for reaching an agreement conditional on the contemporary state x, denoted by $p(x)$, is:

$$p(x) = \Pr\{\epsilon \geq \beta\pi(x) - c_0(x_1) - x_2 | x_1\}. \tag{7}$$

Note the right-hand side of the inequality in (7) is the median of the *potential* loss of the proposer if a unanimous agreement takes place (recall that $c(x, \varepsilon) - \beta\pi(x)$ is the potential gain to the proposer).

Part (iii) states that the dynamics between states can be separated into two Markov chains. This allows for serial correlation in each component in X. This separability holds, for example, if X_1 and X_2 evolve independently over time. A sufficient condition for this is that both X_1 and X_2 follow autoregressive or moving average processes based on independent series of noises. Note that part (iii) can in principle be tested using the time series of observed states in data. Part (iv) is a technical condition that ensures the continuation value, as a solution to certain fixed-point equation, is well-defined and exists.

The first step in our approach is to identify the marginal effect of the additive state X_2 on the continuation value π. This is key for recovering the error distribution, and ultimately the latent surplus. To see why, first note by conditioning on x_1 and varying x_2, we can link the changes in the probability for agreement to the distribution of ϵ given x_1 evaluated at the median potential loss to the proposer. It is clear from (7) that the conditional independence between ϵ and X_2 is instrumental to forming such a link. However, to implement this method we need to know how changes in x_2 affect the potential loss. This is where knowledge of the marginal effect of X_2 on π comes in handy. We need the following assumption in order to recover this marginal effect.

Assumption 3 *$X_2' = \lambda(X_2, \upsilon)$, where the unobserved noise υ is independent from (X_1, X_1') and X_2; λ is increasing in X_2 and υ; and $\partial\lambda(x_2, \upsilon)/\partial x_2 < 1$ for all (x_2, υ).*

Assumption 3 requires that X_2 has limited feedback on X_2'. The orthogonality condition suggests that υ can be considered as serially uncorrelated random noise in the transition. This condition holds, for example, if the series of X_2 follows a stationary first-order autoregressive process: $X_2' = \rho X_2 + \upsilon$, where $|\rho| < 1$ is a constant coefficient and υ is i.i.d. over time and independent from the other observable states in X_1. Note that this condition could be tested using the observed time series of x_2 and x_1 in data.

Lemma 2 *Under Assumptions 1-3, $\pi_2(x) \equiv \partial\pi(x)/\partial x_2$ is identified for all $x \in \mathcal{X}$.*

We provide the intuition for this result in the text, and present the formal proof in Appendix A. For convenience, write the fixed-point characterization of the continuation value in (6) as

$$\pi(x) = \int [\beta\pi(x') + \sigma(x';\pi)]\, dG(x'|x), \tag{8}$$

where

$$\sigma(x;\pi) \equiv \int \max\{c_0(x_1) + x_2 + \varepsilon - \beta\pi(x), 0\} dF(\varepsilon|x_1)$$

is defined as the interim *realized* gains to the proposer in SSPE.

Two insights are crucial for Lemma 2. First, due to the conditional independence between X_2 and ϵ, the marginal effect of X_2 on the interim realized proposer gains $\sigma(X)$ is $p(X)[1 - \beta\pi_2(X)]$, where $p(.)$ is directly identifiable. Second, under Assumption 3, we can identify the distribution of the marginal impact of an additive state on its future value $\partial\lambda(x_2, \upsilon)/\partial x_2$, building on an argument from Matzkin (2003) which exploits the monotonicity of λ in the unobservable noise υ. These two insights, combined with the other restrictions on state transitions in Assumption 2, allow us to use (8) and express the marginal effect of x_2 on π as the fixed point of a contraction mapping whose form depends on identifiable quantities only. This contraction mapping is "quasi-structural" in the sense that it expresses model primitives in terms of identifiable quantities whenever possible. The uniqueness of fixed point in contraction mapping implies π_2 is identified.

We conclude this subsection with an example of an additive state in the context of U.S. corporate bankruptcy cases. When a U.S. firm files for Chapter 11 bankruptcy, it keeps operating while its asset claimants (creditors and equity-holders) negotiate over plans to reorganize the firm. These claimants bargain over the allocation of the firm's expected reorganization value after it emerges from bankruptcy. For this to occur, all asset claimants need to agree on a plan to restructure the firm. The bargaining surplus, or the reorganization value, is determined by factors that affect the firm's market value, and therefore evolves throughout the negotiating process. These factors include firm- or industry-level variables (e.g. stock indexes within certain industrial sector) as well as macroeconomic variables (e.g. inflation rates). The reorganization value may also depend on idiosyncratic noises only known privately to claimants, such as unobserved elements in fixed costs.

Eraslan (2008) compiles a U.S. Corporate Bankruptcy Data (UCBD) which contains 128 large and publicly held firms that filed for Chapter 11 bankruptcy between 1990 and 1997, and that had a confirmed reorganization plan by the end of 2000. For each case filed, the data reports the beginning and the end dates of the negotiation process, as well as firm-level variables that affect the reorganization value. These include the proportion of intangible assets, R&D expenses, and the total claim by creditors at the time of filing. For about 40%

of the firms in data, the reorganization values and the agreed allocations among claimants are not reported. These include cases with and without final agreements, and thus fit in the data scenario we consider in this paper.

In such an environment, a natural candidate for the additive state would be industry-level factors that affect firms' reorganization values. One example is the industry-specific stock index constructed by Merlo and Tang (2012). For each firm in the UCBD, Merlo and Tang (2012) use the Global Industry Classification Standard (GICS) to define its industry category, and use the stock price series in COMPUSTAT to construct a time series of stock indexes pertaining to that industry during the negotiation period (i.e., the interval between the time it files for Chapter 11 and the time its reorganization plan is confirmed).

These industry stock indexes are likely to satisfy the key properties of additive states. First, such indexes reflect sector-level fluctuations in the demand and supplier costs for the whole industry, and therefore affect firms' expected reorganization value. Its additive separability would arise from a popular linear index specification in the surplus function. Second, such industry indexes are constructed by pooling stock prices over a large number of firms in a broad industry category. Thus, their fluctuations contain little information about the evolution of firm-specific factors such as R&D expenses and intangible assets. Third, each firm's idiosyncratic noises are likely to be orthogonal to such an aggregate measure of stock fluctuations on the industry level, at least after controlling for the other firm- or sector-level factors in the data. Fourth, the "limited feedback" condition in Assumption 3 is consistent with a standard specification of autoregressive processes, which are often used in the time series analysis of such stock indexes. Finally, for most industries the indexes demonstrate sufficient large variation over a moderate support.

3.2 Recovering the error distribution

We now show how to recover the error distribution using the marginal effect of X_2 on the continuation value. To do so, we need a mild condition on the scope of variation in X_2.

Assumption 4 *Conditional on any $x_1 \in \mathcal{X}_1$, X_2 is continuously distributed and the support of $p(X)$ includes $\frac{1}{2}$ in the interior.*

This condition states there is sufficient variation in the additive state to induce a moderate probability for unanimous agreement, regardless of the realization of the other states. Given the location normalization in Assumption 2, this is equivalent to the statement that "for all x_1 there exists an x_2 in the interior of the support such that $c_0(x_1) + x_2 = \beta\pi(x_1, x_2)$". This holds if the support of the additive state is large relative to that of the continuation value, which is a continuous and bounded function over $\mathcal{X}$.[3]

[3] Assumption 4 is stronger than necessary for identification. More generally, our argument is valid as long as there exists a known constant $\alpha \in (0, 1)$ that is in the support of $p(X)$ conditional on all $x_1 \in \mathcal{X}_1$. In

The next lemma identifies the error distribution $F_{\epsilon|X_1}$ over parts of its support. For each x_1, let $\underline{\alpha}(x_1)$ and $\overline{\alpha}(x_1)$ denote the infimum and supremum of the support of $1 - p(X)$ conditional on x_1. By definition, both $\underline{\alpha}(x_1)$ and $\overline{\alpha}(x_1)$ are identifiable from the data.

Lemma 3 *Suppose Assumptions 1-4 hold. For all x_1, the α-th conditional quantile of ϵ given x_1 is identified for all $\alpha \in (\underline{\alpha}(x_1), \overline{\alpha}(x_1))$.*

Proof of Lemma 3. For any x_1, let $\xi(x_1) \equiv \inf\{\tilde{x}_2 : p(x_1, \tilde{x}_2) \geq 0.5\}$. By Assumption 4, such an infimum exists and $p(x_1, \xi(x_1)) = 0.5$. Then by construction $c_0(x_1) + \xi(x_1) - \beta\pi(x_1, \xi(x_1)) = 0$. Therefore, for any $x = (x_1, x_2) \in \mathcal{X}$, we can write

$$c_0(x_1) + x_2 - \beta\pi(x) = 0 + \int_{\xi(x_1)}^{x_2} [1 - \beta\pi_2(x_1, \tilde{x}_2)] d\tilde{x}_2, \tag{9}$$

where the right-hand side is identified because of Lemma 2. Also,

$$p(x) = \Pr\{\epsilon \geq \beta\pi(x_1, x_2) - c_0(x_1) - x_2 | x_1\}$$

is directly identifiable from the data. Thus $\beta\pi(x) - c_0(x_1) - x_2$, which is identified above in (9), is the $[1 - p(x)]$-th quantile of ϵ conditional on x_1. $\square$

The lemma recovers the distribution of ϵ given x_1 over the support of the median of the potential loss to the proposer $\beta\pi(X) - c_0(X_1) - X_2$ given x_1. It is not surprising that the distribution of ϵ can only be identified over such a section of the support when there are no additional parametric or shape restrictions. Identification of the error distribution over its full support is still possible if the support of X_2 is large relative to that of the error term and the continuation value.

3.3 Identifying the latent surplus

We present the identification of the latent surplus in two steps. First, we show the latent surplus is identified over its full domain up to an unknown location parameter pertaining to the error distribution. Then, we discuss two alternative approaches for pinning down this location parameter. The first uses the large support of an additive state; the second uses the sufficient exogenous variation in the state transition G or the discount factor β.

To start off, define a location parameter

$$\mu(x_1) \equiv \int_0^\infty \varepsilon dF(\varepsilon | x_1),$$

such a case we can replicate the argument below by replacing the location restriction in Assumption 2 with $F(0|x_1) = \alpha$ for all $x_1 \in \mathcal{X}_1$.

where $F(.|x_1)$ is the distribution of ϵ given x_1. We want to emphasize that $\mu(.)$ is *not* an artifact that can be fixed via normalization without loss of generality.[4] The intuition for the first step of the argument (Lemma 4) is that the ex ante gains to the proposer can be decomposed into $\mu(x_1)$ and an incremental quantity that is identifiable given Lemma 2 and 3. This decomposition helps to recover the continuation value, and consequently c_0, up to the location parameter. To establish Lemma 4, we need the following regularity condition.

Assumption 5 *$\int_0^\infty [1 - F(\varepsilon|x_1)]d\varepsilon < \infty$ for all $x_1 \in \mathcal{X}_1$.*

Assumption 5 restricts the tail behavior of the error distribution. It is used to ensure the existence of some integral that shows up in the proof of identification. It is a mild technical condition satisfied by distributions from several common parametric families.

Lemma 4 *Under Assumptions 1-5, c_0 is over-identified over its full domain $\mathcal{X}_1$ if $\mu(x_1)$ is known for all $x_1 \in \mathcal{X}_1$.*

Proof of Lemma 4. Let $\varphi(x) \equiv \beta\pi(x)$. For any x_1 and $t \in \mathbb{R}$, define $\tilde{\sigma}(x_1, t) \equiv \int \max\{\varepsilon - t, 0\}dF(\varepsilon|x_1)$. By construction $\tilde{\sigma}(x_1, 0) = \mu(x_1)$, and by the Leibniz Rule, $\partial\tilde{\sigma}(x_1, t)/\partial t = -[1 - F(t|x_1)]$ for all $x_1 \in \mathcal{X}_1$. By the second fundamental theorem of calculus,

$$\tilde{\sigma}(x_1, t) = \mu(x_1) - \int_0^t [1 - F(\varepsilon|x_1)]d\varepsilon \tag{10}$$

for all t and x_1. Thus we can write (6) as

$$\begin{aligned} \varphi(x) &= \beta \int [\varphi(x') + \tilde{\sigma}(x'_1, \varphi(x') - c_0(x'_1) - x'_2)]\, dG(x'|x) \qquad (11) \\ &= \beta \int [\varphi(x') + \mu(x_1) - \sigma^*(x)]\, dG(x'|x), \end{aligned}$$

where

$$\sigma^*(x) \equiv \int_0^{\beta\pi(x) - c_0(x_1) - x_2} [1 - F(t|x_1)]dt. \tag{12}$$

The regularity condition in Assumption 5 ensures that the integral in (12) is well-defined for all $x \in \mathcal{X}$. Also note the upper limit of this integral has been identified as in (9). Moreover, by Lemma 3, the error distribution $F(\varepsilon|x_1)$ is identified for all ε in the interval between 0 and $F^{-1}(1 - p(x)|x_1)$ and for all $x \in \mathcal{X}$ conditional on x_1. Hence $\sigma^*(x)$ is identified over the full support $\mathcal{X}$.

Write (11) as

$$\varphi(x) = \beta \int [\varphi(x') - \sigma^*(x')]\, dG(x'|x) + \beta E[\mu(X'_1)|x_1]. \tag{13}$$

[4] This is because our model has already incorporated a location normalization (zero median error) and a scale normalization (marginal effect of additive state on the latent surplus is one). Hence there is no more degrees of freedom for us to restrict the error distribution without loss of generality.

The right-hand side as a functional of φ satisfies the Blackwell Sufficient conditions for contraction mapping; and the form of the functional is identified up to the unknown function $\mu(.)$ because σ^* and G are identified. If $\mu(x_1)$ is known over $\mathcal{X}_1$, then we can recover $\varphi(.)$ over $\mathcal{X}$ as the unique solution to the fixed-point equation (13). Finally, recall that $\varphi(x)-c_0(x_1)-x_2$ is identified for all $x \in \mathcal{X}$ as in (9). Hence $c_0(x_1)$ is (over-)identified over $\mathcal{X}_1$. $\square$

Lemma 4 shows that, to identify the non-linear surplus component c_0, it suffices to recover $\mu(.)$. In the remaining part of this section, we propose two alternatives for doing so. The first is to exploit a large support of X_2 to identify $\mu(.)$ over $\mathcal{X}_1$.

Assumption 6 *Conditional on any x_1, the support of $c_0(x_1) + \epsilon$ is a subset of the support of $\beta\pi(X) - X_2$.*

This support requirement is consistent with the regularity condition (iv) in Assumption 2, which implies the continuation value π is continuous and bounded over $\mathcal{X}$. Under Assumption 6, the support of $p(X)$ given any x_1 is $[0, 1]$. By Lemma 3, this means the distribution of ϵ conditional on X_1 is recovered over its full support, which implies $\mu(x_1)$ is identified for all x_1. The following result then follows immediately from Lemma 4.

Proposition 1 *Under Assumptions 1-6, $c_0(.)$ is identified over its domain $\mathcal{X}_1$.*

The other alternative for recovering $\mu(.)$ from the data is to use exogenous variation is in the transition of states X. To simplify exposition, we focus on the case with discrete states. That is, the support of observed states is finite: $\mathcal{X} \equiv \{x^{(1)}, ., x^{(K)}\}$. It is convenient to adopt vector notation. Let $\mathbf{c}$ and $\boldsymbol{\mu}$ denote two K-vectors with their k-th component being $c_0(x_1^{(k)}) + x_2^{(k)}$ and $\mu(x^{(k)})$ respectively; and let $\mathbf{G}$ be the K-by-K Markov transition matrix between the states, with the (i, j)-th entry being $\Pr(X' = x^{(j)}|X = x^{(i)})$.

The main idea for this second approach is to use variation in state transitions or discount factors across observations of bargaining episodes that share the same latent surplus function. We formalize the notion of sufficient exogenous variation as follows.

Assumption 7 *There exist $M \geq 2$ decision environments, each of which is indexed by $m \leq M$, characterized by a pair of discount factors and transitions $\{\beta^{[m]}, \mathbf{G}^{[m]}\}$ and share the same per-period payoffs $\mathbf{c}$ and error distribution, such that $\mathbf{I}-\beta^{[m]}\mathbf{G}^{[m]}$ has full-rank (K) and*

$$rank\left(\begin{bmatrix} \mathbf{I}-\beta^{[1]}\mathbf{G}^{[1]}, & -\beta^{[1]}\mathbf{G}^{[1]} \\ \vdots & \vdots \\ \mathbf{I}-\beta^{[M]}\mathbf{G}^{[M]}, & -\beta^{[M]}\mathbf{G}^{[M]} \end{bmatrix}_{MK\text{-}by\text{-}2K}\right) = 2K. \tag{14}$$

To implement our idea, a researcher needs to know how to partition the data into subsets defined by distinct environments, where $c(.)$ and $F_{\epsilon|X}$ remain the same while the state transitions change. This is feasible in lots of empirical contexts of dynamic discrete choices. For example, consider a classical example of bus engine replacement in Rust (1987). In this case, the transition of states refers to the evolution of cumulative mileages on a bus engine, which depends on the route travelled by the bus. In each period, the bus manager may choose between engine replacement or maintenance. With engine replacement, cumulative mileages are reset to zero in the next period; with engine maintenance cumulative mileages are carried over to the next period. Then Assumption 7 is plausible because buses traveling different routes may well share the same replacement and maintenance costs (as functions of cumulative mileages). Similar conditions on exogenous variation in decision environments are used in several earlier papers where the distribution of unobserved states are assumed *known*. Examples include Magnac and Thesmar (2002), Fang and Wang (2014) and Norets and Tang (2014).

To see an example of how Assumption 7 holds together with restrictions in part (iii) of Assumption 2, consider the following example with $M = 2$ and both components in the observed states X_1, X_2 being binary random variables. Let $\beta^{[1]} = 0.9$ and $\beta^{[2]} = 0.95$; let the state transitions be $G_1^{[1]} \equiv [.1, .9; .8, .2]$, $G_2^{[1]} \equiv [.3, .7; .6, .4]$, $G_1^{[2]} \equiv [.4, .6; .3, .7]$ and $G_2^{[2]} \equiv [.8, .2; .5, .5]$, where $G_1^{[m]}, G_2^{[m]}$ are transitions of X_1, X_2 in environment m respectively. Assumption 7 holds under such a specification.

Proposition 2 *Suppose $\mathcal{X} \equiv \{x^{(1)}, ..., x^{(K)}\}$ and Assumptions 1-5 and 7 hold. Then $c_0(.)$ and $\mu(.)$ are jointly identified over $\mathcal{X}_1$.*

The main idea behind this proposition is as follows. First off, note that the discounted continuation value can be expressed as the sum of median surplus $c_0(x_1) + x_2$ and the identified quantiles in the error distribution. Thus for each decision environment indexed by m, we can write the Bellman equation characterizing the continuation value in (13) as a system of linear equations that involve the unknown functions $c_0(.)$ and $\mu(.)$. The vector of constants on the right-hand side of this linear system consists of directly identifiable state transitions as well as conditional quantiles of the error and the interim realized gains to the proposer, which are both identified as in the proof of Lemma 4. In matrix notation for discrete states, the coefficients for $\mathbf{c}$ and $\boldsymbol{\mu}$ in such a linear system indexed by m are $\mathbf{I}-\beta^{[m]}\mathbf{G}^{[m]}$ and $-\beta^{[m]}\mathbf{G}^{[m]}$ respectively.

The key for the identification strategy in Proposition 2 is that, under the assumption of exogenous variation, the coefficient matrix and the vector of constants vary across M environments while the unknowns in $\mathbf{c}$ and $\boldsymbol{\mu}$ remain the same. Thus, by stacking equations derived from different decision environments, we effectively add identifying equations without increasing the number of unknowns in the linear system. Joint identification of $\mathbf{c}$ and $\boldsymbol{\mu}$ is therefore achieved provided the exogenous variation in state transition is sufficient in the sense of Assumption 7.

The distribution of the proposer identity L_i does not play any role in our identification analysis. This is due to the "separation principle" that we mention in Section 2. This principle states that, when utility is transferable between the parties, the sum of individual payoffs $w \equiv \sum_i v_i$ and the incidence of a unanimous agreement in a stationary subgame perfect Nash equilibrium depend only on the current states (X, ϵ), but not on the identity of the proposer. The two model elements identified above (i.e., the bargaining surplus and the distribution of unobserved errors) are both linked to the incidence of an agreement only through expected total payoffs $\pi \equiv E[w(X', \epsilon')|x]$, but not the distribution of proposer identities L_i. Furthermore, these two model elements are sufficient for predicting the likelihood of agreement in equilibrium under counterfactual contexts. If the sample data reports the identities of proposers in each round of bargaining, then L_i is immediately identified. Otherwise it is not. Either way it does not affect the identification question or the counterfactual analysis in this paper.

With the identification results established above, one can apply existing estimators for dynamic binary choice models to infer the parameters in the surplus function and the error distribution. To see the link, notice that the characterization of an agreement in our model is analogous to the solution of an optimal stopping problem. One estimator for the dynamic binary choice model is the nested fixed-point maximum likelihood estimator used in Rust (1987) and Rust (1994). Other alternatives include the nested pseudo-likelihood algorithm in Aguirregabiria and Mira (2002) and the asymptotic least squares estimator in Pesendorfer and Schmidt-Dengler (2008). These estimators do require the parametrization of the error distribution. The nonparametric identification result in this section provides a primitive foundation for the conditions required for consistency of these estimators.

4 Extensions

In this section we examine the identification question without the conditions of large support and sufficient exogenous variation in state transitions (Assumptions 6 and 7). In this case, identification depends on whether the error terms are weakly homoskedastic (as defined below), and whether there is exogenous variation in the state transition. In Section 4.1 and 4.2, we show how to recover the surplus up to an unknown location (or equivalently, how to recover the difference in surplus between any pair of states) under different assumptions. In Section 4.3, we show that knowledge of such a location parameter is not necessary for valid counterfactual policy analyses. For the sake of completeness, Section 4.4 concludes by discussing further restrictions that help pin down the location of the surplus.

4.1 Weakly homoskedastic errors

Our first set of results in this section hold when the error terms satisfy a weak form of homoskedasticity as follows.

Assumption 8 *(Weak Homoskedasticity) There exists a constant* $\mu_0 > 0$ *such that* $\mu(x_1) = \mu_0$ *for all* $x_1 \in \mathcal{X}_1$.

Weak homoskedasticity is implied by the stronger restriction that ϵ is independent from X. Now, suppose a researcher recovers the surplus via the following steps. First, calculate the discounted continuation payoff $\varphi(x) \equiv \beta\pi(x)$ by solving the following analog of the fixed-point equation (13):

$$\hat{\varphi}(x) = \beta \int [\hat{\varphi}(x') + \hat{\mu} - \sigma^*(x')] dG(x'|x) \tag{15}$$

where σ^*, defined in (12), is identified over $\mathcal{X}$, and $\hat{\mu}$ is an arbitrarily chosen positive constant to replace the unknown μ_0. Next, recover the non-linear part of surplus as

$$\hat{c}_0(x_1) = \hat{\varphi}(x) - x_2 + \int_{\xi(x_1)}^{x_2} [1 - \beta\pi_2(x_1, \tilde{x}_2)] d\tilde{x}_2 \tag{16}$$

where $\xi(x_1) \equiv \inf\{x_2 : p(x_1, x_2) \geq 0.5\}$ and, as shown in Lemma 2, $\pi_2(x)$ is identified from the probability for agreement $p(x)$ without using $\hat{\mu}$. Then the next proposition shows that $\hat{c}_0$ recovered in this way only differs from the actual c_0 in the data-generating process by a constant.

Proposition 3 *Suppose Assumptions 1-5 and 8 hold. Then for all* x,

$$\hat{c}_0(x_1) = c_0(x_1) + \tfrac{\beta}{1-\beta}(\hat{\mu} - \mu_0).$$

This result is intuitive because if $\mu(x_1) = \mu_0$ does not vary across x_1 in the data-generating process, then the difference between μ_0 and the arbitrarily chosen constant $\hat{\mu}$ does not interact with state transitions in the identification of latent surplus. More specifically, if one uses a standard iterative algorithm to solve for $\hat{\varphi}$ in (15), then the difference $\hat{\mu} - \mu_0$ would be compounded through a geometric series that converges to $\frac{\beta}{1-\beta}(\hat{\mu} - \mu_0)$.

4.2 Heteroskedastic errors with insufficient exogenous variation

Now consider the case where errors are heteroskedastic (so that the weak homoskedasticity in Assumption 8 does not hold), and Assumption 7 fails because there is insufficient variation in state transitions. We show that in this case the surplus function can still be identified up to location under Assumptions 1-5.

We focus on discrete states with matrix notation as before. For example, Assumption 7 fails when exogenous variation only comes from the state transition G but not in the discount factor (that is, $G^{[m]}$ vary but $\beta^{[m]}$ stay the same across $m \leq M$). In this case, the highest rank possible for the MK-by-$2K$ matrix in (14) is $2K-1$ by construction. This is because the sum of the first K columns of the coefficient matrix is proportional to that of the last K columns.

To simplify exposition, suppose $M=2$ and we augment the the system in (14) with an additional location restriction that sets the first component in $\boldsymbol{\mu}$, or $\mu(x^{(1)})$, to an arbitrarily chosen constant $\mu_1^* > 0$. Then solve for $\mathbf{c}$ and (the remaining $K-1$ components in) $\boldsymbol{\mu}$ from the linear system of MK equations that consists of M linear systems as in (26). Denote the solution by $\mathbf{c}^*$ and $\boldsymbol{\mu}^*$. Then

$$\mathbf{c}^* = \mathbf{c} + \mathbf{1}' \tfrac{\beta}{1-\beta} \left[\mu_1^* - \mu\left(x^{(1)}\right)\right],$$

where $\mathbf{c} \in \mathbb{R}^K$ is the actual surplus function in the data-generating process and $\mathbf{1}$ is a K-vector of ones.

The argument is similar to that in Proposition 3. To see why, note that if $(\mathbf{c}, \boldsymbol{\mu})$ is the actual parameter in the data-generating process, then for each $m \leq M$, the structural equation, as shown in (26) in Appendix A, also holds with $(\mathbf{c}, \boldsymbol{\mu})$ replaced by $(\mathbf{c}^*, \boldsymbol{\mu}^*)$, with $\boldsymbol{\mu}^* \equiv \boldsymbol{\mu} + \mathbf{1}' \left[\mu_1^* - \mu\left(x^{(1)}\right)\right]$. This is because the columns in $\mathbf{I} - \boldsymbol{\beta}^{[m]}\mathbf{G}^{[m]}$ (and $\boldsymbol{\beta}^{[m]}\mathbf{G}^{[m]}$) add up to a constant vector of $1-\beta$ (and β respectively) by construction. Thus one can recover the surplus up to an unknown location.

4.3 Counterfactual analyses

We now show that, despite the lack of identification of the location parameter in the surplus, one can still conduct valid counterfactual analyses. Researchers are typically interested in predicting the conditional probability for reaching an agreement in one of the two types of counterfactual experiment:

- *(I):* The surplus function is changed from c to $c + \Delta$, where the change in the surplus function $\boldsymbol{\Delta}$ is known; all other structural elements are unchanged.

- *(II):* The state transition is changed from G to $\bar{G}$, where $\bar{G}$ is known; all other structural elements are unchanged.

We now show that knowledge of the location parameter $\mu(.)$ is not necessary for valid inference of counterfactual outcomes in both cases. For the sake of argument, let the data generating process be the one specified as in Section 4.1. A researcher is interested in predicting the probability for unanimous agreement in the counterfactual context (I). According to theory, a unanimous agreement takes place in SSPE in (I) if and only if $c_0(x_1)+x_2+\Delta(x)+\varepsilon \geq \bar{\varphi}_I(x)$, where $\bar{\varphi}_I(x)$ solves the fixed-point equation

$$\bar{\varphi}_I(x) = \beta \int \left[\bar{\varphi}_I(x') + \mu_0 - \int_0^{\bar{\varphi}_I(x)-c_0(x_1)-x_2-\Delta(x')} [1-F(t|x_1)]dt \right] dG(x'|x). \tag{17}$$

Note the right-hand side of (17) satisfies the Blackwell sufficient conditions for contraction mapping, and therefore the solution $\bar{\varphi}_I$ is unique.

Suppose, in the absence of exogenous variation in state transition, the researcher sets $\mu(x_1)$ to $\hat{\mu}$ for all x_1, and then recovers $\hat{\varphi}$ and $\hat{c}_0$ as in (15) and (16). Then he uses them as inputs to calculate the counterfactual probability for reaching an agreement in SSPE under (I). That is, the probability that $\hat{c}_0(x_1) + x_2 + \Delta(x) + \varepsilon \geq \hat{\varphi}_I(x)$, where $\hat{\varphi}_I(x)$ solves a fixed-point equation that is similar to (17), only with μ_0, c_0 in (17) replaced by $\hat{\mu}, \hat{c}_0$.

The following corollary of Proposition 3 implies that, in such a procedure for counterfactual analyses, the choice of $\hat{\mu}$ has no impact on the predicted probability for reaching an agreement under context (I).

Corollary 1 *Under the conditions stated in Proposition 3, $\hat{\varphi}_I(x) - \hat{c}_0(x_1) = \bar{\varphi}_I(x) - c_0(x_1)$ for all $x \in \mathcal{X}$.*

Recall that the identification of the error distribution $F(\varepsilon|x_1)$ in Lemma 3 only involves the location normalization $F(0|x_1) = 0.5$, but does not rely on any restriction on the location parameter μ_0. Thus the choice of $\hat{\mu}$ has no impact on the conditional error distribution recovered from Lemma 3. It then follows from Corollary 1 that the choice of $\hat{\mu}$ has no impact on the predicted counterfactual probability for agreement $\Pr\{\epsilon \geq \bar{\varphi}_I(x) - c_0(x_1) - x_2 - \Delta(x)|x_1\}$.[5]

Next, we turn to the prediction of agreement probability under the other counterfactual context (II). According to theory, unanimous agreement takes place in SSPE under (II) if and only if $c_0(x_1) + x_2 + \varepsilon \geq \bar{\varphi}_{II}(x)$, where $\bar{\varphi}_{II}(x)$ solves

$$\bar{\varphi}_{II}(x) = \beta \int \bar{\varphi}_{II}(x') + \mu_0 - \left(\int_0^{\bar{\varphi}_{II}(x')-c_0(x_1')-x_2'} [1-F(s|x_1')]ds \right) d\bar{G}(x'|x). \tag{18}$$

Suppose a researcher uses $\hat{c}_0$ as inputs to calculate the probability for reaching an agreement under (II). That is, the probability that $\hat{c}_0(x_1) + x_2 + \varepsilon \geq \hat{\varphi}_{II}(x)$, where $\hat{\varphi}_{II}(x)$ solves a fixed-point equation that is similar to (18), only with μ_0, c_0 in (18) replaced by $\hat{\mu}, \hat{c}_0$. It is straight-forward to verify that $\hat{\varphi}_{II}(x) = \bar{\varphi}_{II}(x) + \frac{\beta}{1-\beta}(\hat{\mu} - \mu_0)$, which implies $\hat{\varphi}_{II}(x) - \hat{c}_0(x_1) =$

[5] It is possible that the correct counterfactual probability $p_\Delta(x) \equiv \Pr\{\epsilon \geq \bar{\varphi}_I(x) - c_0(x_1) - x_2 - \Delta(x)|x_1\}$ is in fact outside the support of $1-p(X)$ given x_1 in the data (i.e., outside the interval $[\underline{\alpha}(x_1), \overline{\alpha}(x_1)]$ defined for Lemma 3). Then in this case one can use our method to partially identify, or bound, the probability $p_\Delta(x)$. The choice of $\hat{\mu}$ has no impact on the bound, or the identified set, calculated in this case.

$\bar{\varphi}_{II}(x) - c_0(x_1)$. Hence by a similar argument the choice of $\hat{\mu}$ does not impact the predicted agreement probability under the context in (II) either.

If the data-generating process is specified as Section 4.2, the same argument can be applied to show that the choice of μ_1^* has no impact on the predicted probability for reaching an agreement under both counterfactual contexts (I) and (II).

4.4 Recovering the location parameter

We conclude this section with discussion about how to identify the location parameter $\mu(.)$ in the error distribution (and the surplus function) without Assumptions 6 and 7. In some cases, the non-linear surplus component may be known at a given state ($c_0(x_1^{(k)}) = \bar{c}$ for a known pair of $\bar{c}$ and $x^{(k)}$). This pins down the location of c_0 and, together with Section 4.1 or 4.2 in this subsection, suggest the surplus function is fully recovered over $\mathcal{X}$.

Alternatively, we could use shape restrictions on the surplus function to achieve the joint identification of μ and c_0. Shape restrictions such as first-degree homogeneity and concavity have been used in the literature to estimate qualitative response models in static settings. (See Matzkin (1994) for details.) Here we apply an argument from Matzkin (1992) to identify the latent surplus under the condition that the non-linear component in the latent surplus is homogeneous of degree one. Specifically, suppose x_1 is a vector and the non-linear surplus component $c_0(x_1)$ is homogenous of degree one. Then one can find a pair of states x and x', where $x \equiv (x_1, x_2)$ and $x' \equiv (x_1', x_2')$, such that $x_1' \equiv \alpha x_1$ for a known $\alpha > 0$ and $\alpha \neq 1$. By degree-one homogeneity, we have

$$c_0(x_1') - c_0(x_1) = (\alpha - 1)c_0(x_1). \tag{19}$$

The left-hand side of (19) is identified because the difference between $c_0(x_1) + x_2$ and $c_0(x_1') + x_2'$ is identified by the results from Section 4.1 or 4.2 and x_2, x_2' are observed states. Hence $c_0(x_1)$ is identified by dividing the left-hand side by $\alpha - 1$. It then follows that c_0 is identified over its full domain $\mathcal{X}_1$.

5 Conclusion

The new identification method we propose in this paper can also be adapted to establish identification in a general optimal stopping model. In a single-agent optimal stopping model, the decision to exit is irreversible. This is analogous to SSB models where the players' decision to reach an agreement is terminal. An optimal stopping problem differs from SSB models in that the agent's payoffs upon exit are typically state-independent, and thus innocuously normalized to a constant (e.g., zero). Therefore, in optimal stopping models, the identification is driven by the variation in additive states that enter per-period payoffs from continuing (as opposed to exiting). Such a discrepancy leads to a different form of Bellman equation. Nevertheless, the intuition for identification in SSB models carries over to the context of optimal stopping. Specifically, under the stated orthogonality conditions on additive states, we can still recover the marginal effect of the additive state on the continuation value as a unique fixed-point in some quasi-structural contraction mapping. Once this is done, the subsequent steps for identifying optimal stopping models follow from arguments similar to those we have provided above for the SSB model. Details are provided in Appendix B.

Appendix A. Proofs

Proof of Lemma 2. Let $\sigma(x)$ denote the interim gains to the proposer. That is,

$$\sigma(x) \equiv \int \max\{c_0(x_1) + x_2 + \varepsilon - \beta\pi(x), 0\} dF(\varepsilon|x).$$

Then under Assumption 2, the partial derivative of σ with respect to x_2 is:

$$\begin{aligned}
\sigma_2(x) &\equiv \frac{\partial}{\partial x_2}\left\{\int_{\beta\pi(x)-c_0(x_1)-x_2} [c_0(x_1) + x_2 - \beta\pi(x) + \varepsilon] dF(\varepsilon|x_1)\right\} \\
&= [1 - \beta\pi_2(x)] \Pr\{\epsilon \geq \beta\pi(x) - c_0(x_1) - x_2 | x_1\} \\
&= [1 - \beta\pi_2(x)] p(x) \qquad (20)
\end{aligned}$$

for all x, where $\pi_2(x) \equiv \partial\pi(x)/\partial x_2$. Thus we can use the right-hand side of (6) to write

$$\begin{aligned}
\pi_2(x) = & \frac{\partial}{\partial x_2}\left\{\int [\beta\pi(x') + \sigma(x')]\, dG(x'|x)\right\} \\
= & \frac{\partial}{\partial x_2}\left\{\int \left(\int [\beta\pi(x_1', \lambda(x_2, \upsilon)) + \sigma(x_1', \lambda(x_2, \upsilon))]\, dG_1(x_1'|x_1)\right) dH(\upsilon)\right\} \qquad (21)
\end{aligned}$$

where H denotes the distribution of υ. Under Assumption 2, we can change the order of differentiation and integration by Lemma 3.6 in Newey and McFadden (1994). Using the equality in (20), we can write the right-hand side of (21) as:

$$\int \lambda_1(x_2, \upsilon) \left(\int [p' + (1 - p')\beta\pi_2']\, dG_1(x_1'|x_1)\right) dH(\upsilon) \qquad (22)$$

where p', π_2' are shorthand for $p(x_1', \lambda(x_2, \upsilon))$ and $\pi_2(x_1', \lambda(x_2, \upsilon))$ respectively and $\lambda_1(x_2, \upsilon) \equiv \partial\lambda(x_2, \upsilon)/\partial x_2$. Then we can write (22) as

$$\int \tilde{\lambda}_1(x_2, x_2') \left(\int [p(x') + (1 - p(x'))\beta\pi_2(x')]\, dG_1(x_1'|x_1)\right) dG_2(x_2'|x_2)$$

where $\tilde{\lambda}_1(x_2, x_2')$ is shorthand for $\lambda_1(x_2, \lambda^{-1}(x_2, x_2'))$ with $\lambda^{-1}(x_2, x_2') \equiv \inf\{\upsilon : \lambda(x_2, \upsilon) \geq x_2'\}$. The equality follows from Assumption 3 using a change of variables between υ and $x_2' = \lambda(x_2, \upsilon)$ while fixing x_2. Thus we have derived a "quasi-structural" characterization of the partial derivative of the continuation value $\pi_2(.)$ as follows:

$$\pi_2(x) = \int \tilde{\lambda}_1(x_2, x_2')\, [p(x') + (1 - p(x'))\beta\pi_2(x')]\, dG(x'|x). \qquad (23)$$

Under Assumption 3, the Blackwell sufficient conditions for contraction mapping hold for the right-hand side. Hence the solution of π_2 in (23) is unique.

Without loss of generality, normalize the distribution of υ to a known distribution. Or equivalently, the α-th quantile of υ, denoted by υ_α, is known for any $\alpha \in (0, 1)$. It then follows that given this normalization, $\lambda(x_2, \upsilon_\alpha)$ is identified for any $\alpha \in (0,1)$ as the α-th quantile of X_2' conditional on x_2 in the data. Hence $\lambda(x_2, \upsilon)$ is identified over its full domain, and so is the partial derivative $\lambda_1(x_2, \upsilon)$. For each pair (x_2, x_2'), the inverse function $\lambda^{-1}(x_2, x_2')$, is identified as the α-th quantile υ_α where $\alpha \equiv \Pr\{X_2' \leq x_2' | x_2\}$. With $\lambda_1(x_2, \upsilon)$ already identified, this implies that $\tilde{\lambda}_1(x_2, x_2')$ is identified. Recall that $p(x')$ and $G(x'|x)$ are directly identifiable from the data. Hence we can recover $\pi_2(.)$ as the unique solution to (23) over the full support of X. $\square$

Proof of Proposition 2. For each environment m, let $\boldsymbol{\varphi}^{[m]}$ denote a K-by-1 vector with the k-th component being the discounted continuation value in an environment indexed by m. We can then rewrite (13) in matrix notation as:

$$\begin{aligned} & \boldsymbol{\varphi}^{[m]} = \beta^{[m]} \mathbf{G}^{[m]} \left(\boldsymbol{\varphi}^{[m]} - \boldsymbol{\sigma}^{[m]} \right) + \beta^{[m]} \mathbf{G}^{[m]} \boldsymbol{\mu} \\ \Rightarrow \quad & \boldsymbol{\varphi}^{[m]} = (\mathbf{I} - \beta^{[m]} \mathbf{G}^{[m]})^{-1} \beta^{[m]} \mathbf{G}^{[m]} \left(\boldsymbol{\mu} - \boldsymbol{\sigma}^{[m]} \right) \end{aligned} \tag{24}$$

where $\boldsymbol{\sigma}^{[m]}$ is a K-by-1 vector with the k-th component being $\sigma^*(x^{(k)})$ in environment m, and $\mathbf{Q}^{[m]}$ denotes a K-by-1 vector with the k-th entry being the $p_{k,m}$-th quantile of the error distribution, where $p_{k,m}$ equals the probability for *not* reaching an agreement under the state $x^{(k)}$ in an environment indexed by m. Recall (from the proof of Lemma 4) that the k-th component in $\boldsymbol{\sigma}^{[m]}$ is identified as $\int_0^{F^{-1}\left(p_{k,m}|x^{(k)}\right)} \left[1 - F\left(\varepsilon | x^{(k)}\right)\right] d\varepsilon$, where $F^{-1}(.|x)$ is the quantile function of ϵ conditional on x.

Write (7) in matrix notation as

$$\boldsymbol{\varphi}^{[m]} - \mathbf{c} = \mathbf{Q}^{[m]}. \tag{25}$$

Under Assumptions 1-4, Lemma 3 implies that $\mathbf{Q}^{[m]}$ is identified for all $m \leq M$. Substitute (25) into (24) to get:

$$(\mathbf{I} - \boldsymbol{\beta}^{[m]} \mathbf{G}^{[m]}) \mathbf{c} - \boldsymbol{\beta}^{[m]} \mathbf{G}^{[m]} \boldsymbol{\mu} = -(\mathbf{I} - \boldsymbol{\beta}^{[m]} \mathbf{G}^{[m]}) \mathbf{Q}^{[m]} - \boldsymbol{\beta}^{[m]} \mathbf{G}^{[m]} \boldsymbol{\sigma}^{[m]}. \tag{26}$$

Recall that $\mathbf{G}^{[m]}$ directly identifiable from the data; and $\beta^{[m]}$ is assumed known. Also, $\mathbf{Q}^{[m]}, \boldsymbol{\sigma}^{[m]}$ are identified as discussed above. Hence the right-hand side of (26) is identified and so are the coefficients for $\mathbf{c}$ and $\boldsymbol{\mu}$ on the left-hand side of (26). Stacking such equations across M environment, we get a linear system of MK equations and $2K$ unknowns (in $\mathbf{c}$ and $\boldsymbol{\mu}$). Under Assumption 7, the linear system admits a unique solution in $\boldsymbol{\mu}$ and $\mathbf{c}$. This implies $\mu(.)$ and $c_0(.)$ are over-identified over the support of X_1. $\square$

Proof of Proposition 3. Under Assumption 8, the fixed-point equation that characterizes the discounted continuation value $\varphi(x) \equiv \beta\pi(x)$, or equation (13), is reduced to

$$\varphi(x) = \beta \int \left[\varphi(x') - \sigma^*(x')\right] dG(x'|x) + \beta\mu_0$$

where σ^* is defined as in (12). Let $\hat{\varphi}$ be defined as in the text. Then by construction

$$\hat{\varphi}(x) - \varphi(x) = \beta \int [\hat{\varphi}(x') - \varphi(x')] dG(x'|x) + \beta (\hat{\mu} - \mu_0).$$

By the regularity condition in Assumption 5, both φ and $\hat{\varphi}$ are continuous and bounded over the support of X. Then recursive substitution implies that

$$\hat{\varphi}(x) - \varphi(x) = \lim_{J\to\infty} \sum_{j=1}^{J} \beta^j (\hat{\mu} - \mu_0) + \lim_{J\to\infty} \beta^J \int [\hat{\varphi}(x') - \varphi(x')] dG(x'|x), \tag{27}$$

where the first term equals $\frac{\beta}{1-\beta}(\hat{\mu} - \mu_0)$ and the second term converges to 0 for all x because the integral is bounded for all x under Assumption 2. Recall from the proof of Lemma 3 that

$$c_0(x_1) = \varphi(x) - x_2 + \int_{\xi(x_1)}^{x_2} [1 - \beta\pi_2(x_1, \tilde{x}_2)] d\tilde{x}_2 \tag{28}$$

where $\xi(x_1)$ and π_2 are both identified from the observed probability for agreement $p(x)$ in the data without using the choice of $\hat{\mu}$. It then follows from (16), (27) and (28) that

$$\hat{c}_0(x_1) - c_0(x_1) = \hat{\varphi}(x) - \varphi(x) = \tfrac{\beta}{1-\beta}(\hat{\mu} - \mu_0).$$

This proves the claim in the proposition. $\square$

Proof of Corollary 1. By Proposition 3,

$$\hat{c}_0(x_1) = c_0(x_1) + \tfrac{\beta}{1-\beta}(\hat{\mu} - \mu_0)$$

where $\mu_0 = \mu(x_1)$ is the actual location parameter in the data-generating process. As a result, if $\bar{\varphi}_I(x)$ is the unique solution to (17), then

$$\hat{\varphi}_I(x) = \bar{\varphi}_I(x) + \tfrac{\beta}{1-\beta}(\hat{\mu} - \mu_0)$$

is the unique solution to

$$\hat{\varphi}_I(x) = \beta \int \hat{\varphi}_I(x') + \hat{\mu} - \left(\int_0^{\hat{\varphi}_I(x') - \hat{c}_0(x_1') - x_2' - \Delta(x')} [1 - F(s|x_1')] ds \right) dG(x'|x).$$

Thus $\hat{\varphi}_I(x) - \hat{c}_0(x_1) = \bar{\varphi}_I(x) - c_0(x_1)$. $\square$

Appendix B. Extension to Optimal Stopping Models

In this section we show how to adapt the method in the text to answer the same identification questions in general optimal stopping models. Consider a single-agent optimal stopping problem where the decision to exit is irreversible and hence terminal. Upon exit ($D = 0$), the agent receives a state-independent payoff that is normalized to zero. The per-period payoff from continuing ($D = 1$) is $u_o(x_1) + x_2 + \varepsilon$, where the state (x, ε) evolves according to a first-order Markov process that satisfies conditional independence in Assumption 1.

Let the additive state x_2 satisfy orthogonality conditions in Assumption 2. That is, ε is independent from x_2 given x_1 with $F(0|x_1) = 0.5$ and the state transition is multiplicatively separable $G(x'|x) = G_1(x_1'|x_1)G_2(x_2'|x_2)$. Also suppose the evolution in x_2 is governed by $x_2' = \lambda_o(x_2, \eta)$ where η is independent from (x_2, x_1, x_1'), and λ_o is increasing in both arguments with $\partial\lambda_o/\partial x_2 < 1$ over the full domain of λ_o. The optimal decision is to continue if and only if $v_o(x) + \varepsilon \geq 0$, where $v_o(.)$ solves

$$v_o(x) = u_o(x_1) + x_2 + \beta \int \sigma_o(x'; F, v_o) dG(x'|x), \tag{29}$$

where $\sigma_o(x; F, v_o) \equiv \int \max\{v_o(x) + \varepsilon, 0\} dF(\varepsilon|x_1)$ is the agent's interim gains from continuation.

Marginal effect of additive states on the value function. As in SSB models, the first step for identification is to recover the marginal effect of the additive state x_2 on the deterministic part of the value function v_o. Similar to Section 3, our strategy is to using the exogenous variation in x_2 to construct a fixed-point characterization for this marginal effect. Note that by an application of the Leibniz rule,

$$\partial\sigma_o(x)/\partial x_2 = v_{o,2}(x)p_o(x), \tag{30}$$

where $v_{o,2}(x) \equiv \partial v_o(x)/\partial x_2$ and $p_o(x) \equiv \Pr\{v_o(x) + \varepsilon \geq 0|x\}$. Let H_o denote the distribution of η, and let $\lambda_{o,1}(x_2, \eta) \equiv \partial\lambda_o(x_2, \eta)/\partial x_2$. Under the maintained assumptions on the additive states, we can write the marginal effect of x_2 on the integral in the right-hand side of (29) as

$$\begin{aligned}
& \tfrac{\partial}{\partial x_2} \int \sigma_o(x_1', \lambda_o(x_2, \eta); F, v_o) dG(x_1'|x_1) dH(\eta) \\
= & \int \lambda_{o,1}(x_2, \eta) \int v_{o,2}(x_1', \lambda_o(x_2, \eta)) p_o(x_1', \lambda_o(x_2, \eta)) dG(x_1'|x_1) dH_o(\eta) \\
= & \int \tilde{\lambda}_{o,1}(x_2, x_2') p_o(x') v_{o,2}(x') dG(x'|x)
\end{aligned} \tag{31}$$

where $\tilde{\lambda}_{o,1}(x_2, x_2')$ is shorthand for $\lambda_{o,1}(x_2, \lambda_o^{-1}(x_2, x_2'))$ with $\lambda_o^{-1}(x_2, x_2') \equiv \inf\{\eta : \lambda_0(x_2, \eta) \geq x_2'\}$. The first equality in the display above follows from changing the order of integration and differentiation and (30); the second from changing variables between η and $x_2' = \lambda_o(x_2, \eta)$ while fixing x_2.

Differentiating both sides of (29) with respect to x_2 and substituting in (31), we get a "quasi-structural" fixed-point characterization of the marginal effect $v_{o,2}$:

$$v_{o,2}(x) = 1 + \beta \int \tilde{\lambda}_{o,1}(x_2, x_2')p_o(x')v_{o,2}(x')dG(x'|x). \tag{32}$$

Under the assumption that $\partial\lambda_o/\partial x_2 < 1$, the Blackwell sufficient conditions for contraction mapping hold for the right-hand side of (32). Hence the solution of $v_{o,2}$ in (32) is unique. By the same argument as in the last paragraph in the proof of Lemma 2, $\tilde{\lambda}_{o,1}(x_2, x_2')$ is identified over its full domain. With β assumed known and $p_o(.)$ directly identified from the data, this means the marginal effect $v_{o,2}$ is identified as the unique solution to the fixed-point equation (32).[6] With $v_{o,2}(x)$ identified as above, we can adapt the remaining steps in Section 3 and 4 in the optimal stopping model.

Value function and the distribution of unobserved states. Suppose the support condition in Assumption 4 holds. That is, conditional on any x_1, X_2 is continuously distributed and the support of $p(X)$ given x_1 includes $\frac{1}{2}$ in the interior. Then for any x_1, let $\xi_o(x_1) \equiv \inf\{\tilde{x}_2 : p_o(x_1, \tilde{x}_2) \geq 0.5\}$. Under the support condition, this infimum exists for all x_1 and $p_o(x_1, \xi_o(x_1)) = 0.5$. Then by construction $v_o(x_1, \xi_o(x_1)) = 0$. For any x, we can write

$$v_o(x) = 0 + \int_{\xi_o(x_1)}^{x_2} v_{o,2}(x_1, \tilde{x}_2)d\tilde{x}_2.$$

The right-hand side is identified because both ξ_o and $v_{o,2}$ are identified. This means for all x, the $[1 - p_o(x)]$-th quantile of ε conditional on x_1 (which equals $-v_o(x)$) is identified.

Surplus process. For any x_1 and $t \in \mathbb{R}$, define $\tilde{\sigma}_o(x_1, t; F) \equiv \int \max\{t + \varepsilon, 0\}dF(\varepsilon|x_1)$. Define $\mu(x_1) \equiv \tilde{\sigma}_o(x_1, 0; F)$. By the Leibniz Rule, $\partial\tilde{\sigma}(x_1, t; F)/\partial t = 1 - F(-t|x_1)$ for all x_1. Thus by the second fundamental theorem of calculus,

$$\tilde{\sigma}_o(x_1, t; F) = \mu(x_1) + \int_0^t [1 - F(-\varepsilon|x_1)]d\varepsilon$$

for all t and x_1. We can write (6) as

$$\begin{aligned} v_o(x) &= u_o(x_1) + x_2 + \beta \int \tilde{\sigma}_o(x_1', v_o(x'); F)dG(x'|x) \\ &= u_o(x_1) + x_2 + \beta \int [\mu(x_1) + \sigma_o^*(x)]\, dG(x'|x), \end{aligned}$$

where

$$\sigma^*(x) \equiv \int_0^{v_o(x)} [1 - F(-t|x_1)]dt.$$

[6] We are grateful to Victor Aguirregabiria for showing the marginal effect of an additive state x_2 on the value function can be identified in an optimal stopping model when x_2 evolves according to an additive structure $x_2' = \lambda(x_2) + \eta$.

By our argument above, the mild support condition in Assumption 4 implies that for all x, the upper limit of this integral $v_o(x)$ is identified and the integrand $1-F(-t|x_1)$ is identified for all $t \in [-v_o(x), 0]$. It then follows that $\sigma^*(x)$ is identified for all x. This establishes a counterpart of Lemma 4 in the context of optimal stopping models. That is, if $\mu(x_1)$ is known, then $u_o(x_1)$ will be (over-)identified for all x_1 as

$$u_o(x_1) = v_o(x) - x_2 - \beta \int [\mu(x_1) + \sigma_o^*(x)]\, dG(x'|x).$$

Location, counterfactual, etc. Given the results above, it is straightforward to adapt the remaining steps (Proposition 1 and 2) in Section 3.3 and the extended results in Section 4 in the context of optimal stopping.

REFERENCES

Aguirregabiria, V. and P. Mira. 2002. "Swapping the Nested Fixed Point Algorithm: A Class of Estimators for Discrete Markov Decision Models," Econometrica, vol. 70, pp. 1519-1543.

Benjamin, D. and M. Wright. 2009. "Recovery Before Redemption? A Theory of Delays in Sovereign Debt Renegotiations," Working Paper, UCLA

Berry, S. and P. Haile. 2014. "Identification in Differentiated Products Markets Using Market Level Data," Econometrica, vol. 82, pp. 1749-1798.

Bi, R. 2008. "'Beneficial' Delays in Debt Restructuring Negotiations," Working Paper 08/38, IMF

Chen, S., Khan, S. and X. Tang. 2016. "Informational Content of Special Regressors in Heteroskedastic Binary Response Models," Journal of Econometrics, Vol. 193, pp 162-182

Diaz-Moreno, C. and J. Galdon-Sanchez. 2004. "Collective Bargaining Under Complete Information," in: Salomon W. Polachek (Ed.), Accounting for Worker Well-Being, Research in Labor Economics, 2004, Vol. 23, pp. 359-379.

Diermeier, D., Eraslan, H. and A. Merlo. 2003. "A Structural Model of Government Formation," Econometrica, Vol. 71, pp. 27-70.

Eraslan, H. 2008. "Corporate Bankruptcy Reorganizations: Estimates from a Bargaining Model," International Economic Review, Vol. 49, pp. 659-681.

Fang, H. and Y. Wang. 2014. "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting, with an Application to Mammography Decisions," International Economic Review, forthcoming

Ghosal, S., Miller, M. and K. Thampanishvong. 2010. "Delay and Haircuts in Sovereign Debt: Recovery and Sustainability," mimeo, University of Warwick.

Lewbel, A. 2000. "Semiparametric Qualitative Response Model Estimation With Unknown Heteroscedasticity and Instrumental Variables," Journal of Econometrics, Vol. 97, pp. 145-177

Lewbel, A. and X. Tang. 2015. "Identification and Estimation of Games with Incomplete Information Using Excluded Regressors," Journal of Econometrics, Vol. 189, pp. 229-244

Magnac, T. and D. Thesmar. 2002. "Identifying Dynamic Discrete Decision Processes," Econometrica, Vol. 70, pp. 801-816.

Matzkin, R.L. 1992. "Nonparametric and Distribution-Free Estimation of the Binary Choice and the Threshold Crossing Models", Econometrica, Vol. 60, pp. 239.

Matzkin, R.L. 1994. "Restrictions of Economic Theory in Nonparametric Methods," Handbook of Econometrics, Vol. 4, edited by C.F. Engel and D.L. McFadden, Elsevier.

Matzkin, R.L. 2003. "Nonparametric Estimation of Nonadditive Random Functions," Econometrica, Vol. 71, pp. 1339-1375.

Merlo, A. 1997. "Bargaining over Governments in a Stochastic Environment," Journal of Political Economy, Vol. 105, pp. 101-131.

Merlo, A. and X. Tang. 2012. "Identification and Estimation of Stochastic Sequential Bargaining Games," Econometrica, Vol. 80, pp. 1563-1604

Merlo, A. and C. Wilson. 1995. "A Stochastic Model of Sequential Bargaining with Complete Information," Econometrica, Vol. 63, p. 371-399.

Merlo, A. and C. Wilson. 1998. "Efficient Delays in a Stochastic Model of Bargaining," Economic Theory, Vol. 11, pp. 39-55.

Norets, A. and X. Tang. 2014. "Semiparametric Inference in Dynamic Binary Choice Models," Review of Economic Studies, Vol. 81, pp. 1229-1262

Pesendorfer, M. and P. Schmidt-Dengler. 2008. "Asymptotic Least Squares Estimators for Dynamic Games," Review of Economic Studies, Vol. 75, pp. 901-928.

Rust, J. 1987. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," Econometrica, Vol. 55, pp. 999-1033.

Rust, J. 1994. "Structural Estimation of Markov Decision Processes," Handbook of Econometrics, Vol. 4, edited by R. Engle and D. McFadden, Elsevier, North Holland.

Simcoe, T. 2012. "Standard Setting Committees: Consensus Governance for Shared Technology Platforms," American Economic Review, Vol. 102, pp. 305-36.