# A Comment on "Estimating Dynamic Discrete Choice Models with Hyperbolic Discounting" by Hanming Fang and Yang Wang[*]

Jaap H. Abbring[†]        Øystein Daljord[‡]

May 2019

**Abstract**

Fang and Wang (2015)'s Proposition 2 claims generic identification of a dynamic discrete choice model with hyperbolic discounting under exclusion restrictions. We note that Proposition 2 uses a definition of "generic" that does not preclude that a generically identified model is nowhere identified. We provide two examples of models that are generically identified under this definition, but that are, respectively, everywhere and nowhere identified. We then show that its proof is incorrect and incomplete. We conclude that Proposition 2 has no implications for identification of the dynamic discrete choice model and suggest alternative approaches to identification.

---

# 1 Introduction

Fang and Wang (2015) studied the identification an infinite-horizon, stationary dynamic discrete choice model with partially naive hyperbolic time preferences and is frequently cited in the recent dynamic discrete choice literature (e.g. Yao et al., 2012; Lee, 2013; Ching et al., 2013; Norets and Tang, 2014; Dubé et al., 2014; Gordon and Sun, 2015; Bajari et al., 2016; Chan, 2017; Gayle et al., 2018). In its model, the agent chooses an action $i$ from $\mathcal{I} \equiv \{0, 1, \ldots, I\}$, $I \in \mathbb{N}$, in each period, after she observes that period's Markov state $(x, \varepsilon)$, where $x$ takes values in a finite set $\mathcal{X}$ and $\varepsilon = (\varepsilon_0, \varepsilon_1, \ldots, \varepsilon_I) \in \mathbb{R}^{I+1}$.[1] This returns instantaneous utility $u_i^*(x, \varepsilon) = u_i(x) + \epsilon_i$. It also controls the evolution of $x$: Given choice $i$ in state $(x, \epsilon)$, it takes the value $x' \in \mathcal{X}$ in the next period with probability $\pi(x'|x, i)$. In contrast, the components of $\varepsilon'$ are mutually independent with type-1 extreme value distributions, independently from $x'$, $(x, \varepsilon)$, and choice $i$. The agent has rational expectations; in particular, she believes $x$ to evolve according to the controlled Markov transition distribution $\pi$. She discounts future utility with a standard factor $\delta$ and present bias factor $\beta$, and perceives future selves to have present bias factor $\tilde{\beta}$. With a normalization $u_0(x) = 0$ for all $x \in \mathcal{X}$, the model's unknown primitives are an $IX$-vector $\mathbf{u}$ with the values of $u_i(x)$ for $i \in \mathcal{I}/\{0\}$ and $x \in \mathcal{X}$, the discount function parameters $(\beta, \tilde{\beta}, \delta)$, and a matrix $\mathbf{\Pi}$ with the state transition probabilities $\pi(x'|x, i)$ for $i \in \mathcal{I}$, $x \in \mathcal{X}$, and $x' \in \mathcal{X}$. Here, $X = |\mathcal{X}|$ is the number of elements of $\mathcal{X}$.

The econometrician's data are the state transition probabilities $\mathbf{\Pi}$ and a matrix $\tilde{\mathbf{P}}$ that collects the conditional probabilities $P_i(x)$ that the agents chooses $i$ in state $x$, $i \in \mathcal{I}$ and $x \in \mathcal{X}$. Because probabilities sum to one, the $(I+1)X + (I+1)X^2$ choice and transition probabilities in $\left(\tilde{\mathbf{P}}, \mathbf{\Pi}\right)$ can be represented by a vector that stacks $IX + (I+1)X(X-1)$ of them. We adopt this representation and take $\left(\tilde{\mathbf{P}}, \mathbf{\Pi}\right) \in [0, 1]^{IX+(I+1)X(X-1)} \subset \mathbb{R}^{IX+(I+1)X(X-1)}$.[2]

---

[1]This paper's footnotes document various minor errors and inconsistencies in Fang and Wang (2015) that are not central to our comments, but that we have corrected in the main text to ensure clarity and consistency. Here, for example, we have included $\varepsilon_0$ in $\varepsilon$. Fang and Wang (2015) (p. 568) specified $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_I) \in \mathbb{R}^I$ and only assumed $u_i^*(x, \varepsilon) = u_i(x) + \epsilon_i$ for $i \in \mathcal{I}/\{0\}$. However, Fang and Wang (2015) subsequently used $u_0^*(x, \varepsilon) = u_0(x) + \varepsilon_0$, with $\varepsilon_0, \ldots, \varepsilon_I$ independent with type-1 extreme value distributions, to get logit choice probabilities $P_i(x)$.

[2]Fang and Wang (2015)'s online Appendix C instead specifies

$$\left(\tilde{\mathbf{P}}, \mathbf{\Pi}\right) \in \Delta^{(I+1)X} \times \left(\overbrace{\Delta^X \times \cdots \times \Delta^X}^{X \text{ copies}}\right)^{I+1} \subset \mathbb{R}^{IX+(I+1)X(X-1)},$$

without defining $\Delta$. We guess that, for $J \in \mathbb{N}$, $\Delta^J \equiv \{(p_1, \ldots, p_J) \in \mathbb{R}^J : p_1 \geq 0, \ldots, p_J \geq 0; \sum_{j=1}^J p_j = 1\}$ denotes a probability simplex, but then $\Delta^{(I+1)X} \times \left(\Delta^X \times \cdots \times \Delta^X\right)^{I+1}$ lies in a $IX + (I+1)X(X-1)$-dimensional linear subspace of $\mathbb{R}^{(I+1)X+(I+1)X^2}$ rather than in $\mathbb{R}^{IX+(I+1)X(X-1)}$. However, all that matters for the reading of Fang and Wang (2015)'s Proposition 2 and our comments is that Fang and Wang (2015) use Lebesgue measure on $\mathbb{R}^{IX+(I+1)X(X-1)}$ to decide between generic and exceptional sets of data; see Footnote 10.

The transition probabilities $\mathbf{\Pi}$ directly identify the agent's (rational) beliefs. The conditional choice probabilities $\tilde{\mathbf{P}}$ are linked to the model's primitives by the assumption that the agent's actions follow a *stationary perception-perfection perfect strategy profile* of the decision problem with beliefs $\mathbf{\Pi}$ and some utilities $\mathbf{u}^*$ and discount factors $(\beta^*, \tilde{\beta}^*, \delta^*)$. The extreme-value assumption ensures that the conditional choice probabilities $P_i(x)$ have the logit form. As in the special case with geometric discounting ($\beta = \tilde{\beta} = 1$), an application of Hotz and Miller (1993)'s choice probability inversion gives $IX$ equations that relate the $IX + 3$ parameters $(\mathbf{u}, \beta, \tilde{\beta}, \delta)$ to the data $(\tilde{\mathbf{P}}, \mathbf{\Pi})$, one for each log choice probability contrast $\ln P_i(x) - \ln P_0(x)$, $i \in \mathcal{I}/\{0\}$ and $x \in \mathcal{X}$.[3]

For its main identification result (Proposition 2), Fang and Wang (2015) (p. 579) assumed that the observed state can be partitioned as $x = (x_r, x_e)$; where $x_r$ takes values in $\mathcal{X}_r$, $x_e$ takes values in $\mathcal{X}_e$, $\mathcal{X} = \mathcal{X}_r \times \mathcal{X}_e$, and $|\mathcal{X}_e| \geq 2$; and that its Assumption 5 holds for all $(x_r, x_e) \in \mathcal{X}$ and $(x_r, x_e') \in \mathcal{X}$.[4] The first part[5] of its Assumption 5 then requires that

$$u_i(x_r, x_e) = u_i(x_r, x_e') \text{ for all } i \in \mathcal{I}/\{0\}, \ x_r \in \mathcal{X}_r, \text{ and } (x_e, x_e') \in \mathcal{X}_e \times \mathcal{X}_e, \tag{1}$$

which are $I(|\mathcal{X}_e| - 1)|\mathcal{X}_r|$ different and nontrivial exclusion restrictions, one for each $i \in \mathcal{I}/\{0\}$, each $x_r \in \mathcal{X}_r$, and each of the $|\mathcal{X}_e| - 1$ distinct pairs of subsequent $x_e$ and $x_e'$ in the (arbitrarily) ordered set $\mathcal{X}_e$.[6]

Taken together, the $IX$ constraints resulting from Hotz and Miller (1993)'s choice probability inversion and those in (1) implied by the exclusion restrictions form a system of $IX + I(|\mathcal{X}_e| - 1)|\mathcal{X}_r|$ nonlinear equations in the $IX + 3$ parameters $(\mathbf{u}, \beta, \tilde{\beta}, \delta)$ and the data $(\tilde{\mathbf{P}}, \mathbf{\Pi})$. Fang and Wang (2015) denoted this system with

$$\tilde{\mathcal{G}}\left(\mathbf{u}, \beta, \tilde{\beta}, \delta; \left(\tilde{\mathbf{P}}, \mathbf{\Pi}\right)\right) = 0. \tag{2}$$

The system of equations (2) contains all the information linking the unknown parameters $(\mathbf{u}, \beta, \tilde{\beta}, \delta)$ to the data $(\tilde{\mathbf{P}}, \mathbf{\Pi})$ under the assumed exclusion restrictions in (1). Therefore,

---

[3]Fang and Wang (2015) provided the analysis leading to these equations, but not the final equations themselves.

[4]Fang and Wang (2015) used the same notation for random variables and their realizations and, in Assumption 5, incorrectly referred to the state's values $x_1$ and $x_2$ as "state variables."

[5]The second part of Assumption 5 requires that transition probabilities $\pi(\cdot|x, i)$ for some choice $i$ differ between the same $(x_r, x_e)$ and $(x_r, x_e')$. This condition cannot possibly be necessary for Fang and Wang (2015)'s Proposition 2 to be true, as it holds generically according to its definition of "generic" (see Section 2).

[6]Fang and Wang (2015)'s online Appendix C instead states that "the data must also satisfy the additional $I \times |\mathcal{X}_e| \times |\mathcal{X}_r|$ equations requiring that $u_i(x_r, x_e) = u_i(x_r)$ for each $i \in \mathcal{I}/\{0\}$, each $x_e \in \mathcal{X}_e$ and each $x_r \in \mathcal{X}_r$." Its subsequent analysis fails to appreciate that these $I \times |\mathcal{X}_e| \times |\mathcal{X}_r|$ equations come with $I \times |\mathcal{X}_r|$ additional parameters $u_i(x_r)$; $i \in \mathcal{I}/\{0\}$, $x_r \in \mathcal{X}_r$ (it concludes that the exclusion restrictions yield a system of "$I \times X + I \times |\mathcal{X}_e| \times |\mathcal{X}_r|$ equations in $I \times X + 3$ unknowns $(\mathbf{u}, \beta, \tilde{\beta}, \delta)$"). Clearly, on balance, these $I|\mathcal{X}_e||\mathcal{X}_r|$ equations only introduce $I|\mathcal{X}_e||\mathcal{X}_r| - I|\mathcal{X}_r| = I(|\mathcal{X}_e| - 1)|\mathcal{X}_r|$ additional restrictions, as in our representation. Of course, these restrictions are simply the equalities in (1) that can be derived by differencing Fang and Wang (2015)'s equations $u_i(x_r, x_e) = u_i(x_r)$ and $u_i(x_r, x_e') = u_i(x_r)$.

Fang and Wang (2015) studied their model's identification by analyzing whether (2) uniquely determines $(\mathbf{u}, \beta, \tilde{\beta}, \delta)$ for given data. It claimed the following result:[7]

> **Proposition 2.** Consider the space of data sets that can be generated by the assumed data generating process for some primitives $(\mathbf{u}^*, \beta^*, \tilde{\beta}^*, \delta^*)$. Suppose that there exist state variables that satisfy Assumption 5. Then, all the model parameters are generically identified if $I(|\mathcal{X}_e| - 1)|\mathcal{X}_r| \geq 4$.

Fang and Wang (2015) does not formally define "generic identification," but paraphrases Proposition 2 as giving identification "for almost all data sets generated by the assumed hyperbolic discounting model" (p. 579). Fang and Wang (2015)'s proof of Proposition 2, in its online Appendix C, further defines "almost all" and therewith "generic."

The proof of Proposition 2 applies the transversality theorem to Fang and Wang (2015)'s model to demonstrate that there are generically no parameters that are consistent with any given data. Next, it notes that since the model generated the data by assumption, there must exist some parameters consistent with the data. It concludes that these parameters are therefore generically the only parameters that are consistent with such data.

Section 2 uncovers Fang and Wang (2015)'s definition of "generic" from this proof. Then, it demonstrates that Proposition 2 is *void*: The model may be generically identified, in the sense of Fang and Wang (2015), independently of whether any data sets that can be generated by the model correspond to a unique parameter vector. It then shows that the proof is *incorrect*. Finally, it notes that the proof is *incomplete*, as it fails to verify the rank condition for the transversality theorem that it invokes. It is shown that independently of whether this rank condition holds, the proof has no implications for the model's identification. Section 3 concludes with a brief discussion of alternative approaches to identification in dynamic discrete choice models.

## 2  A void generic identification result

The proof of Proposition 2 in Fang and Wang (2015)'s online Appendix C first presents the following transversality theorem (Mas-Colell, 1985, Proposition 8.3.1):[8]

**Theorem 1.** Let $F : A \times B \to \mathbb{R}^m$, $A \subset \mathbb{R}^n$, $B \subset \mathbb{R}^s$ be $C^r$ with $r > \max\{n - m, 0\}$. Suppose that 0 is a regular value of $F$; that is, $F(a, b) = 0$ implies rank $\partial F(a, b) = m$. Then, except for a set of $b \in B$ of Lebesgue measure zero, $F_b : A \to \mathbb{R}^m$ has 0 as a regular value.

---

[7]We quote Fang and Wang (2015)'s Proposition 2 verbatim, except that we have replaced its condition $I|\mathcal{X}_e||\mathcal{X}_r| \geq 4$ with the stronger condition $I(|\mathcal{X}_e| - 1)|\mathcal{X}_r| \geq 4$. Proposition 2's proof in Fang and Wang (2015)'s online Appendix C relies on the fact that "$I \times X + I \times |\mathcal{X}_e| \times |\mathcal{X}_r|$ ... is larger than the number of unknowns $I \times X + 3$ under our identifying assumption that $I \times |\mathcal{X}_e| \times |\mathcal{X}_r| \geq 4$." However, as we have explained in Footnote 6, the number of equations equals $IX + I(|\mathcal{X}_e| - 1)|\mathcal{X}_r|$, not $IX + I|\mathcal{X}_e||\mathcal{X}_r|$, so that $I(|\mathcal{X}_e| - 1)|\mathcal{X}_r| \geq 4$ is required to ensure that there are more equations than unknowns. Note that this correction neither changes the substance of Fang and Wang (2015)'s proof, which simply relies on having more equations than unknowns, nor that of our comment.

[8]To avoid confusion with Fang and Wang (2015)'s use of $x$ for states, we slightly deviate from Mas-Colell (1985)'s and Fang and Wang (2015)'s notation and use $a$ instead of $x$ and $A$ instead of $N$ here.

Here, $\partial F$ is the Jacobian of $F$ with respect to $(a, b)$; $F_b$ is such that $F_b(a) = F(a, b)$ for all $a, b$; and $\partial F_b$ is the Jacobian with respect to $a$ only.

To prove Proposition 2, it applies this transversality theorem to the system of equations (2), with the following mapping of notation:[9]

| Transversality Theorem | Fang and Wang (2015) |
|---|---|
| $F(a, b) \in \mathbb{R}^m$ | $\tilde{\mathcal{G}}\left(\mathbf{u}, \beta, \tilde{\beta}, \delta; \left(\tilde{\mathbf{P}}, \mathbf{\Pi}\right)\right) \in \mathbb{R}^{IX + I(|\mathcal{X}_e|-1)|\mathcal{X}_r|}$ |
| $m$ | Number of equations in $\tilde{\mathcal{G}}$: $IX + I\left(|\mathcal{X}_e| - 1\right)|\mathcal{X}_r|$ |
| $a \in A \subset \mathbb{R}^n$ | Unknown parameters $\left(\mathbf{u}, \beta, \tilde{\beta}, \delta\right) \in \mathbb{R}^{IX} \times (0, 1]^3 \subset \mathbb{R}^{IX+3}$ |
| $n$ | Number of unknown parameters: $IX + 3$ |
| $b \in B \subset \mathbb{R}^s$ | Vector of probabilities in $[0, 1]^{IX+(I+1)X(X-1)} \subset \mathbb{R}^{IX+(I+1)X(X-1)}$ |
| | that represents the data $\left(\tilde{\mathbf{P}}, \mathbf{\Pi}\right)$ |
| $s$ | $IX + (I+1)X(X-1)$ |
| $a \mapsto F_b(a)$ | $\left(\mathbf{u}, \beta, \tilde{\beta}, \delta\right) \mapsto \tilde{\mathcal{G}}\left(\mathbf{u}, \beta, \tilde{\beta}, \delta; \left(\tilde{\mathbf{P}}, \mathbf{\Pi}\right)\right)$ |

That is, Fang and Wang (2015) studied the generic identification of the vector $a \in A \subset \mathbb{R}^n$ of unknown parameters $\left(\mathbf{u}, \beta, \tilde{\beta}, \delta\right)$ from the choice and transition probabilities $b \in B \subset \mathbb{R}^s$ by applying the transversality theorem to the system $F(a, b) = 0$ of $m$ smooth equality constraints. This implicitly defines "for almost all data sets generated by the assumed hyperbolic discounting model" (and therewith "generically" in Proposition 2) to mean for all data $b \in B$ in the model's range (the set of $b \in B$ such that $F(a, b) = 0$ has at least one solution $a \in A$) outside a set of Lebesgue measure zero in $\mathbb{R}^s$.[10]

A key problem with Fang and Wang (2015)'s Proposition 2 is that its proof, given that the assumed rank condition holds (we return to this at the end of this section), establishes that the model's range has Lebesgue measure zero in $\mathbb{R}^s$. Because the transversality theorem, as applied in Fang and Wang (2015)'s proof, only has implications for data outside a set of Lebesgue measure zero, it has no consequences for identification from data in the model's range.

To be precise, suppose that the rank condition for the transversality theorem holds: rank $\partial F(a, b) = m$ if $F(a, b) = 0$. Then, the transversality theorem implies that, for all $b \in B$ outside a set of Lebesgue measure zero in $\mathbb{R}^s$, rank $\partial F_b(a) = m$ if $F(a, b) = 0$. Moreover, because $a \in \mathbb{R}^n$, rank $\partial F_b(a) \leq n < m$. Taken together, this implies that $F(a, b) = 0$ has no solutions $a \in A$, except for $b \in B$ in a set of Lebesgue measure zero in $\mathbb{R}^s$.

---

[9]This mapping corrects two minor problems with Fang and Wang (2015)'s mapping at the top of page 3 of its online Appendix. See Footnotes 2 and 6.

[10]Recall from Footnote 2 that it is not completely clear how Fang and Wang (2015) represent the choice and transition probability data, but that it is clear that they think of the data as living in $\mathbb{R}^s = \mathbb{R}^{IX+(I+1)X(X-1)}$. The exact way the data are represented in $\mathbb{R}^s$ is irrelevant, because Lebesgue measure is invariant under affine transformations with determinant 1 or $-1$. In particular, Fang and Wang (2015)'s representation and ours both assign zero measure to the same sets of choice and transition probabilities.

Consequently, given that the rank condition holds, Proposition 2 is void. It claims that $F(a, b) = 0$ has a unique solution $a \in A$ for all $b \in B$ in the model's range, but since the model's range has zero measure, it is excepted from the claim. Proposition 2 therefore makes no claim about the number of solutions in the range of the model. We note that Proposition 2 is not false. Formally, Proposition 2 is vacuously true, because it is a statement about a property of the elements of an empty set.[11]

Moreover, its proof cannot easily be adapted to establish a more substantive identification result, for some or all data in the model's range, because its application of the transversality theorem has no implications for the number of parameters $a \in A$ that solves $F(a, b) = 0$ for data $b \in B$ in the model's zero measure range.

We illustrate these two points with two simple examples. We first note that Fang and Wang (2015)'s proof does not use the particular structure of the dynamic discrete choice model, but applies to any model that can be represented by a system of equations with more equations than unknowns under the regularity conditions stated above. Our examples therefore use highly stylized, linear models that allow easy and direct verification of the rank condition and the conclusions of the transversality theorem. Like Fang and Wang (2015)'s model under the conditions of Proposition 2, both examples have models with more equations than unknowns ($m > n$). Their ranges have Lebesgue measure zero in $\mathbb{R}^s$, so that generic identification vacuously holds. However, in the first example, $a$ is uniquely determined from $F(a, b) = 0$ for $b \in B$ in the model's range; in the second example, it is not.

**Example 1 (Everywhere point identified).** Suppose that the data are $b = (b_1, b_2) \in B = \mathbb{R}^2$, the parameter is $a \in A = \mathbb{R}$, and the model is $F : \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$, with

$$0 = F(a; b) = \begin{bmatrix} b_1 - a \\ b_2 - a \end{bmatrix}, \quad \partial F(a; b) = \begin{bmatrix} -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}, \text{ and } \partial F_b(a) = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

Note that $n = 1$, $s = 2$, and $m = 2$. In this example, rank $\partial F(a, b) = 2$ always. The transversality theorem gives that $F_b(a) = 0$ implies rank $\partial F_b(a) = 2$ for almost all $b \in \mathbb{R}^2$. Now, rank $\partial F_b(a) \leq n = 1$, so $F_b(a) \neq 0$ for almost all $b \in \mathbb{R}^2$. The model is linear, so we can do without the transversality theorem and directly observe that the model can only generate data $b$ such that $b_1 = b_2$, which is nongeneric in $B = \mathbb{R}^2$. Data $b \in B$ that can be generated by this model uniquely determine $a$. In this example, the transversality theorem tells us that there are zero (not one) parameters that rationalize the data, for almost all data in $\mathbb{R}^2$, and that data that are in the model's range, excepted from the transversality theorem, *always* identify the unknown parameter.

**Example 2 (Nowhere point identified).** Now suppose we have data $b = (b_1, b_2, b_3) \in B = \mathbb{R}^3$, a pair of parameters $a = (a_1, a_2) \in A = \mathbb{R}^2$, and a model is $F : \mathbb{R}^2 \times \mathbb{R}^3 \to \mathbb{R}^3$, with

$$0 = F(a; b) = \begin{bmatrix} b_1 - (a_1 + a_2) \\ b_2 - (a_1 + a_2) \\ b_3 - (a_1 + a_2) \end{bmatrix},$$

---

[11]It is vacuously true since any statement about a property of elements of an empty set is formally true.

$$\partial F(a;b) = \begin{bmatrix} -1 & -1 & 1 & 0 & 0 \\ -1 & -1 & 0 & 1 & 0 \\ -1 & -1 & 0 & 0 & 1 \end{bmatrix}, \text{ and } \partial F_b(a) = \begin{bmatrix} -1 & -1 \\ -1 & -1 \\ -1 & -1 \end{bmatrix}.$$

Note that $n = 2$, $s = 3$, and $m = 3$. In this example, rank $\partial F(a, b) = 3$ always. The transversality theorem gives that $F_b(a) = 0$ implies that rank $\partial F_b(a) = 3$ for almost all $b \in \mathbb{R}^3$. Since rank $\partial F_b(a) \leq n = 2$, $F_b(a) \neq 0$ for almost all $b$. This again makes sense: The model, which requires $F(a, b) = 0$, can only generate data $b$ such that $b_1 = b_2 = b_3$, which is nongeneric in $\mathbb{R}^3$. Data from the range of the model only identify $a_1 + a_2$ and never $a_1$ and $a_2$ separately. So, this is another example where transversality tells us there are zero (not one) parameters that match the data for almost all data. However, unlike in the previous example, data in the model's range *never* identify the parameters.

Together, these examples show that the transversality theorem, as applied in Fang and Wang (2015)'s proof, has no implications for identification. Given that the rank condition for its application of the transversality condition holds, Fang and Wang (2015)'s proof is correct up to its last half sentence. The first half of the proof's last sentence correctly concludes that, for all data $b \in B$ outside a set of Lebesgue measure zero, there exist no parameters $a \in A$ that solve $F(a, b) = 0$. However, the last half sentence qualifies this conclusion with "except the true primitives $(\mathbf{u}^*, \beta^*, \tilde{\beta}^*, \delta^*)$... that generated the data." This qualification does not follow from the preceding mathematical arguments. In particular, we have shown that Fang and Wang (2015)'s application of the transversality theorem implies that zero, not one, parameters solve the model for all data outside a set of measure zero. The transversality theorem has no implications for the number of parameters that solve the model for data in an exceptional set, which includes the model's range. So, this last half sentence of Fang and Wang (2015)'s proof is incorrect.

Finally, Fang and Wang (2015)'s proof is also incomplete, because it fails to verify the key rank condition for its application of the transversality theorem: rank $\partial F(a, b) = m$ if $F(a, b) = 0$. Instead, Fang and Wang (2015) noted that "this can be verified in the same way that we verify [a similar condition in the proof of Proposition 1]," but did not verify the latter condition either (p.578). The incompleteness of the proof is however immaterial for the conclusion that can be drawn from the proof. If the rank condition holds, we know that the model's range has Lebesgue measure zero and is excepted from the transversality theorem. If the rank condition is violated, the transversality theorem does not apply. Either way, the proof has no implications for identification.

# 3 Discussion

The source of the problems with Fang and Wang (2015)'s Proposition 2 is its focus on identification that is generic in the data space, rather than the parameter space. This is nonstandard and complicates the analysis in two ways. First, the specification of an appropriate measure directly on the data requires knowledge of the model's empirical content, i.e. the range of data that can be generated by varying the model parameters on their domain.

Our discussion of Proposition 2 highlights the problems of ignoring the model's empirical content.

Second, it is unclear how the concept of generic identification in the data space corresponds to the concept of generic identification in the parameter space. The two concepts are generally not interchangeable, as the following stylized example illustrates. Consider a model that maps a parameter $\theta \in \mathbb{R}$ to a choice probability $p = p(\theta) \in [0, 1]$. Define "for almost all" $\theta$ (or $p$) to mean for all $\theta$ (or $p$) outside a set of Lebesgue measure 0. If $p(\theta) = 1/(1 + \exp(\theta))$, then $\theta$ is identified for almost all $p$ and almost all $\theta$. If instead $p(\theta)$ equals 0 if $\theta \leq 0$, $\theta$ if $\theta \in (0, 1)$, and 1 if $\theta \geq 1$, then $\theta$ is identified for almost all $p$, but is not identified for almost all $\theta$.

One could possibly derive an identification result for the case with more equations than free parameters that is generic in the parameter space instead, following e.g. Sargan (1983), McManus (1992), and Ekeland et al. (2004). One would also have to choose between a measure-theoretic definition of "genericity", like Fang and Wang (2015)'s, and a topological one. McManus (1992) and Ekeland et al. (2004) provide discussion. Generic identification, however, is a weak concept of identification, and particularly so if the exceptional set cannot be characterized. The very small subsets where identification fails may happen to contain economically important models. One example is Ekeland et al. (2004) which shows that the generic identification of the hedonic model does not cover the linear-quadratic special case that is at the center of most applied work.

Abbring and Daljord (2019) offers an alternative approach that dispenses with the concept of generic identification. It instead exploits the specific structure of the dynamic discrete choice model to analyze identification of a special case of Fang and Wang (2015)'s model, with exponential discounting ($\beta = \tilde{\beta} = 1$). It shows that each exclusion restriction in (1) for some $x_r \in \mathcal{X}_r$ and distinct $x_e \in \mathcal{X}_e$ and $x'_e \in \mathcal{X}_e$ gives a single moment condition that relates the discount factor $\delta$ to the choice and transition probabilities.[12] These moment conditions contain all the information in the data about $\delta$, and therewith $\mathbf{u}$.[13] The analysis shows that each single exclusion restriction in general gives set identification, where the identified set is finite. For important special cases, such as models with one-period finite dependence (e.g. Rust, 1987; De Groote and Verboven, 2019), the exclusion restriction gives point identification. Abbring et al. (2019) showed that it is similarly possible to concentrate the identification analysis of a model with sophisticated present biased preferences ($\tilde{\beta} = 1$) on a small number of moment conditions derived directly from equally many exclusion restrictions. However, this approach does not extend to Fang and Wang (2015)'s partially naive case. The identification of partially naive time preferences seems to require an analysis of the full system of equations and remains an open question.

---

[12]The exclusion restrictions in (1) are special cases of the ones considered in Abbring and Daljord (2019).

[13]Abbring and Daljord (2019)'s Section 4 noted that a version of Magnac and Thesmar (2002)'s Proposition 2 holds: There exist unique (up to a standard utility normalization) values of the primitives (notably, $\mathbf{u}$) that rationalize the data for any given discount factor $\beta \in [0, 1)$. The joint identification of a non-parametric utility function and the discount factor is therefore reduced to the conditions on $\beta$ derived from the exclusion restriction.

# References

Abbring, J. H. and Ø. Daljord (2019). Identifying the discount factor in dynamic discrete choice models. Working Paper 2018–55 (revised), Becker Friedman Institute, Chicago. arXiv:1808.10651 [econ.EM].

Abbring, J. H., Ø. Daljord, and F. Iskhakov (2019). Identifying present-biased discount functions in dynamic discrete choice models. Mimeo, CentER, Tilburg.

Bajari, P., C. S. Chu, D. Nekipelov, and M. Park (2016). Identification and semiparametric estimation of a finite horizon dynamic discrete choice model with a terminating action. *Quantitative Marketing and Economics 14*(4), 271–323.

Chan, M. K. (2017, March). Welfare dependence and self-control: An empirical analysis. *Review of Economic Studies 84*(4), 1379–1423.

Ching, A., T. Erdem, and M. P. Keane (2013). Learning models: An assessment of progress, challenges, and new developments. *Marketing Science 32*(6), 931–938.

De Groote, O. and F. Verboven (2019, June). Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems. *American Economic Review 109*(6), 2137–2172.

Dubé, J.-P., G. J. Hitsch, and P. Jindal (2014). The joint identification of utility and discount functions from stated choice data: An application to durable goods adoption. *Quantitative Marketing and Economics 12*(4), 331–377.

Ekeland, I., J. J. Heckman, and L. Nesheim (2004). Identification and estimation of hedonic models. *Journal of Political Economy 112*(1), S60–S109.

Fang, H. and Y. Wang (2015). Estimating dynamic discrete choice models with hyperbolic discounting, with an application to mammography decisions. *International Economic Review 56*(2), 565–596.

Gayle, G.-L., L. Golan, and M. A. Soytas (2018). Estimation of dynastic life-cycle discrete choice models. *Quantitative Economics 9*(3), 1195–1241.

Gordon, B. R. and B. Sun (2015, May-June). A dynamic model of rational addiction: Evaluating cigarette taxes. *Marketing Science 34*(3), 452–470.

Hotz, V. J. and R. A. Miller (1993). Conditional choice probabilities and the estimation of dynamic models. *Review of Economic Studies 60*(3), 497–529.

Lee, R. (2013). Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review 103*(7), 2960–3000.

Magnac, T. and D. Thesmar (2002). Identifying dynamic discrete choice processes. *Econometrica 70*, 801–816.

Mas-Colell, A. (1985, October). *The Theory of General Economic Equilibrium.* Number 9780521265140 in Cambridge Books. Cambridge University Press.

McManus, D. A. (1992). How common is identification in parametric models? *Journal of Econometrics 53*(1), 5–23.

Norets, A. and X. Tang (2014). Semiparametric inference in dynamic binary choice models. *Review of Economic Studies 81*(3), 1229–1262.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica 55*, 999–1033.

Sargan, J. D. (1983). Identification and lack of identification. *Econometrica 51*(6), 1605–1633.

Yao, S., C. F. Mela, J. Chiang, and Y. Chen (2012). Determining consumers discount rates with field studies. *Journal of Marketing Research 49*(6), 822–841.