Identification of Dynamic Panel Binary Response Models*

S. Khan[†] M. Ponomareva[‡] E. Tamer[§]

March 2019

Abstract

We analyze identification in dynamic econometric models of binary choice with fixed effects under general conditions. This class of models is often used in the literature to distinguish between state dependence (invariably referred to in the recent literature as switching costs, inertia or stickiness) and heterogeneity. We fir7(fi)6(r)2pac.6(e)4(t)0(e)4(r)4(i)z.4(c)4fi6(t)9(4fi6t

the same exercise under the stronger assumption of *conditional exchangeability*, and establish its incremental identifying power. We extend our identification approach to study models with more time periods as well. We also provide sufficient conditions for point identification. For inference in cases with discrete regressors, we provide an approach to constructing confidence sets for the identified sets using a linear program that is simple to implement. The paper then provides simulation based evidence on the size and shape of the identified sets in varying designs to illustrate the informational content of different assumptions. We also illustrate the inference approach using a data set on women's labor supply decisions.

Keywords: Binary Choice, Dynamic Panel Data, Partial Identification.

JEL: C22, C23, C25.

*We are grateful for helpful comments from S. Chen, A. Chesher, H.Ichimura, T. Magnac, as well as conference and seminar participantsch,elan,
Toulouse, UCL, UCLA, University of Connecticut, and Wuhan. X. Lin, P. Sarkis and Z. Wang provided excellent research assistance.

[†]Department of Economics, Boston College, Chestnut Hill, MA, shakeeb.khan@bc.edu

[‡]Department of Economics, Northern Illinois University, DeKalb, IL, mponomareva@niu.edu

[§]Department of Economics, Harvard University, Cambridge, MA, elietamer@fas.harvard.edu

1 Introduction

There has been recent renewed interest in empirical economics in estimating models of discrete choice over time. This is partly motivated by empirical regularities: certain individuals are more likely to stay with a choice if they have experienced that choice in the past and this choice "stickiness" has been attributed variably in the literature to *inertia* or *switching costs*. For example, Handel (2013) estimates a model of health insurance choice in a large firm over time documenting *inertia* in choices overtime. Dubé, Hitsch, and Rossi (2010) empirically find that this "inertia" in packaged goods markets is likely caused by brand loyalty. Polyakova (2016) studies the important question of quantifying the e□ect of switching costs in Medicare Part D markets and its relation to adversely selected plans[1]. The recent availability of these panel data in such important markets on the one hand and the central role that the dynamic discrete choice literature played in econometric theory on the other provide the main motivation for this paper which is focused on the question of identification in these models.

The dynamic discrete choice model has appeared prominently in econometrics. In fundamental work, Heckman (1981) discusses two di□erent explanations for the empirical regularity that an individual is more likely to experience a state after having experienced it in the past. The first explanation, termed *state dependence*, is a genuine behavioral response to occupying the state in the past, i.e., a similar individual who did not experience the state in the past is less likely to experience it now. The current literature sometimes refers to state dependence as switching costs, inertia or stickiness and can be thought of as a *causal e ect* of past occupancy of the state[2]. The second explanation advanced by Heckman is *heterogeneity*, whereby individuals are di□erent in unobservable ways and if these unobservables are correlated over time, this will lead to said regularity. This serial correlation in the

unob-344527ularit unob-3445270.0006T-31he1Tc11.9559552223.127560.614Tm/Ty114leunob-1(at)(o)50(ge

Concretely, the binary dynamic panel data model relates a binary outcome in period t, y_t (we abstract from subscripting also by i) to its lagged value y_{t-1}

the sign of γ in a model with and without covariates with $T = 2, 3$. Complementing our are proven to have informational content in the sense that they result in smaller identified regions than the model in Section 3.1, yet still more general than the models introduced in Chamberlain (1985) and Honoré and Kyriazidou (2000). Section 4 considers extensions of the model to allow for a panel data with a longer time series. Specifically, in 4.1 we add additional time periods to the panel, exploring the time component's informational content by showing how the identified region shrinks when more periods are available. Section 5 compliments our identification results in the previous sections by proposing computationally attractive methods to conduct inference on the structural parameters. This will enable testing, for example, if there is indeed *persistence* in the binary variable of interest. Section 6 explores the finite sample properties of our procedures with an empirical application on female employment status as well as reporting results from simulation studies which explore how the identified region varies across the different models considered. Section 7 concludes with discussions on areas for future research, such as the effect of introducing more choices available to the agent, by studying a dynamic multinomial choice model with individual and choice effects, as first introduced in Chamberlain (1984) and more recently in Pakes and Porter (2014) and Ouyang, Khan, and Tamer (2017).

2 Dynamic Panel Binary Choice Model

Recall our model of the form:

$$y_t = I\{u_t \leq x_t\beta + \gamma y_{t-1} + \alpha\} \tag{2.1}$$

where u_t is an unobserved scalar random variable, x_t is an observed $k-$dimensional vector of covariates, β denotes an unknown k dimensional vector of regression coefficients, α denotes the unobserved scalar individual specific effect. The observed binary variable y_t takes the value 1 if the argument inside the indicator function $I\{\cdot\}$ is true, and 0 otherwise. Finally, we let the unknown scalar parameter γ denote the measure of persistence in the model.

In what follows we will will explore the identifiability of the unknown parameters β, γ, when making one of the following assumptions about the distribution of $u_1, u_2, \ldots, u_T$:

(STAT)

we assume that n is large relative to

3.2 Exchangeability with $T = 2$

In this section, we replace the conditional (on α and x) stationarity assumption with conditional (on α, x, y_0) *exchangeability*[6] of idiosyncratic error terms and investigate its identifying power.

Definition 1. *A sequence* $u_1, u_2, \ldots, u_T$

$u_{-M}, \ldots, u_0$, $x_{-M}, \ldots, x_0$, and α; and at the same time y_0 (the first observable outcome) is a *deterministic* function of that history.

Given the above, we analyze the identifying power of the conditional exchangeability of idiosyncratic error terms u_t's, summarized in the following assumption:

Assumption 3.2. *(CEX): $u_1, \ldots, u_T$ are exchangeable conditional on α, x, y_0.*

An even stronger alternative to the stationarity Assumption 3.1

is interesting since the result in Proposition 3.1 provides a characterization of the identified set in the Manski model without any conditions[7] on the support of x.

The next Theorem, whose proof follows, is the main result in this section.

Theorem 3.2. *Suppose that Assumption 2.1 holds. Let $\Theta_{I,cex}^{\{1,2\}}(2)$ be the set of parameters that satisfy conditions (1) and (2) of Proposition 3.1. Also let $\Theta_{I,cex}^{\{1,2\}}(1)$ satisfy the following restriction: if for some $z = (x, y_0)$*

(1) $P(y_1 = 1|z) \geq P(y_2 = 1|z) \Longrightarrow (x_2 - x_1)\tilde{\beta} + \min\{0, \tilde{\gamma}\} - \tilde{\gamma}y_0 \leq 0$;

(2) $P(y$

521.793 Tm/Ty131T

(1) $P(y_1 = 1, y_2 = 0|x, y_0) \geq P(y_1 = 0, y$

Proof: Although Lemma 3.2 requires us to look at only some sequences $(a_1, \ldots, a_T)$ and to match only the distribution of indicator variables $y_1, \ldots, y_T$

model

$$\tilde{y}_t = I\{\tilde{u}_t \leq x_t\tilde{\beta} + \tilde{\gamma}\tilde{y}_{t-1} + \tilde{\alpha},\ t = 1, 2\}$$

where the distribution $\tilde{F}_{\tilde{u},\tilde{\alpha}|z}$ obeys the exchangeability assumption, and where the distribution of $(\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, x$

matches $P(y_1 = 0, y_2 = 1|z)$.

right of

$$\text{where } \delta_1 + \delta_2 + \delta_3 = 1 \text{ and } d_j \geq 0$$

A bivariate Fréchet copula is symmetric. If the joint distribution of $\tilde{u}_1$ and $\tilde{u_2}$ is defined by a symmetric copula, i.e. if

$$\tilde{F}_{\tilde{u}_1,\tilde{u}_2|z} = \tilde{C}(\tilde{F}, \tilde{F})$$

then $\tilde{u}_1$ and $\tilde{u}_2$ are exchangeable.

Given a solution $q_1(z)$, $q_{20}(z)$ and q

Konstantopoulos and Yuan (2018)). Finally, Lemma 3.2 guarantees that there exists $\tilde{F}_{\tilde{u},\tilde{\alpha}|z}$ such that $\tilde{u}_1, \ldots, \tilde{u}_t$ are iid conditional on $\tilde{\alpha}$ and z, and by construction,

$$p(\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, x|\tilde{\theta}, \tilde{F}_{\tilde{u},\tilde{\alpha}|z}) = p(y_0, y_1, y_2, x|\theta, F_{u,\alpha|z})$$

□

Finally, to complete the proof of sharpness of $\Theta_{I,cex}^{\{1,2\}}$, we need to show that θ is identified relative to any $\tilde{\theta}$ / $\Theta_{I,cex}^{\{1,2\}}$ under either conditional exchangeability or conditional iid assumptions. Assume that for a given $\tilde{\theta}$ / $\Theta_{I,cex}^{\{1,2\}}$ e.g. condition (1) of Theorem 3.2 does not hold. That is, there exists some $z = (y_0, x)$ such that

$$P(y_1 = 1|z) \geq P(y_2|z) \text{ and } (x_2 - x_1)\tilde{\beta} + \min\{0, \tilde{\gamma}\} - \tilde{\gamma}y_0 > 0$$

However, if there exists $\tilde{F}_{\tilde{u},\tilde{\alpha}|z}$ F_{cex} such that

$$p(\tilde{y}_0, \tilde{y}_1, \tilde{y}_2, x|\tilde{\theta}, \tilde{F}_{\tilde{u},\tilde{\alpha}|z}) = p(y_0, y_1, y_2, x|\theta, F_{u,\alpha|z})$$

then it must be the case that (see the proof of Lemma 3.3)

$(x_2 - x_1)\tilde{\beta}$ + min{0, /Ty241Tf()TjETBT11.9552421.4350Tm350.983Tm/Ty1

(1) If $P(y_1 = 1, y_2 = 0|x, y_0) \geq 1 - P(y_1 = 0, y_2 = 1|\tilde{x}, \tilde{y}_0)$, then $((x_1 - x_2) - (\tilde{x}_1 - \tilde{x}_2))\beta + \gamma(y_0 - \tilde{y}_0 - 1) \geq 0$.

(2) If $P(y_1 = 1, y_2 = 0|x, y_0) > 1 - P(y_1 = 0, y_2 = 1|\tilde{x}, \tilde{y}_0)$, then $((x_1 - x_2) - (\tilde{x}_1 - \tilde{x}_2))\beta + \gamma(y_0 - \tilde{y}_0 - 1) > 0$.

Remark 3.1. *Note that the independence condition of Proposition 3.2 keeps the independence assumption from Honoré and Kyriazidou (2000) but relaxes stationarity.*

Given that $y_0, \tilde{y}_0$ are binary, by looking at (1) and (2) in the Proposition 3.2 above, we see that only the sign of γ will be identified, but we may get some meaningful identification for β. We can also potentially add more restrictions to shrink the identified set even further. For example, if we assume that $Med(u_1 - u_2|x, y_0) = 0$, then the following must hold for any x, y_0 in the support:

(1) If $P(y_1 = 1, y_2 = 0|x, y_0) \geq 0.5$, then $(x_1 - x_2)\beta + \gamma(y_0 - 1) \geq 0$.

(2) If $P(y_1 = 0, y_2 = 1|x, y_0) \geq 0.5$, then (

Theorem 3.1

where for $j = 1, 2, 9, 10$*:*

$$m_j = \inf_{\Delta x \quad \Delta X_j} \Delta x\,\beta, \; M_j = \sup_{\Delta x \quad \Delta X_j} \Delta x\,\beta$$

.

Note that the identification of the sign of γ in this result does not rely on β being point identified. However, when the sign of γ is identified, we can weaken Assumption 3.4. In particular, if γ is positive, then we can replace X_7 and X_8 in Assumption 3.4 with $X_3 \quad X_7$ and X_4

with additional conditions on observed regressors. Again, we define the following sets:

$\Delta X_1(y_0$

(2) If $\Delta X_2(0) \quad \Delta X_4(1) = \varnothing$ *or* $\Delta X_3(0) \cap \Delta X_1(1) = \varnothing$, *then* $\gamma >$

(11) $P(y_2$

Remark 4.1. *When* $x_2 = x_3$, *restrictions for event pairs* $\{(0,1,0),(1,0,0)\}$ *and* $\{(0,1,1),(1,0,1)\}$ *in Table 3 reduce to the following set:*

(1) If $(x_1 - x_2)\beta + \gamma(y_0 - 1) \geq 0$, *then* $p_3(1,0,1|\alpha, x, y_0) \geq p_3(0,1,1|\alpha, x, y_0)$

(2) If $(x_1 - x_2)\beta + \gamma(y_0 - 1) \leq 0$, *then* $p_3(1,0,1|\alpha, x, y_0) \leq p_3(0,1,1|\alpha, x, y_0)$

(3) If $(x_1 - x_2)\beta + \gamma y$

(6) $P(y_2 = 1, y_3 = 0|x, y_0) \geq P(y_1 = 0|x, y_0)$ $\quad (x_3 - x_1)\beta + \gamma - \gamma y_0 \leq 0.$

Finally, let $\Theta_{I,cex}^{\{2,3\}}(1)$ *satisfy the restrictions for* $\Theta_{I,stat}^{\{2,3\}}$ *in Theorem 4.1, only with the conditional on x probabilities replaced by the conditional on* $z = (x, y_0)$ *probabilities. Then* $\Theta_{I,cex}^{T=3} = \Theta_{I,cex}^{\{1,2,3\}} \cap \Theta_{I,cex}^{\{1,2\}}(1) \cap \Theta_{I,cex}^{\{2,3\}}(1) \cap \Theta_{I,cex}^{\{1,3\}}(1)$ *is the sharp identified set for* θ *under either Assumption 3.2 (exchangeability) or Assumption 3.3 (conditional independence).*

Here the intersection of sets $\Theta_{I,stat}^{\{1,2\}}$, $\Theta_{I,stat}^{\{2,3\}}$ and $\Theta_{I,stat}^{\{1,3\}}$ gives us the set of parameters that are observationally equivalent to the true parameter under the assumption that u_1, u_2 and u_3 are identically distributed conditional on x, y_0, and α. Set $\Theta_{I,cex}^{\{1,2,3\}}$ gives us the set of parameters that are observationally equivalent to the true parameter under conditional (on x, y_0, and α) exchangeability of u_1, u_2, u_3. Note that unlike the case with $T = 2$, some of the exchangeability restrictions are not implied but any of the stationarity restrictions. For example, exchangeability-based restriction 1.a in Table 3 is a stronger version the following stationarity-based restriction for $\Theta_{I,cex}^{\{2,3\}}(1)$:

$$P(y_1 = 0, y_2 = 0|x, y_0) \geq P(y_3 = 0|x, y_0) \qquad (x_3 - x_2)\beta + \max\{0, \gamma\} \geq 0$$

and so on.

4.2 Point Identification with $T = 3$

In this section we provide sufficient conditions for point identification of the parameters β, γ under the stationarity and exchangeability assumptions in the case for $T = 3$. Point identification under stationarity will rely on the result in Theorem 4.1, while point identification under exchangeability will be based on Theorem 4.2.

We start with point identification under stationarity. Theorem 4.1 provides 6 conditions that involve β only, so we can use these conditions to point identify β in a similar way we did for $T = 2$. In particular, we define the following sets

$$X_7^{\{1,2\}} = \{x \quad X \text{ such that } P(y$$

stationarity restriction (i.e. condition (9) in 4.2 holds), then the sign of γ is also be identified under the exchangeability restriction (condition (3) in 4.3 holds), but the reverse is not true.

When $T = 3$, it is sometimes possible to identify the sign of γ even when γ is positive (unlike in $T = 2$ case). We start with stationarity Assumption 3.1: under that assumption, the sharp identified set for β and γ is given by Theorem 4.1. In the absence of covariates, this result (in addition to the restrictions on γ described above for $T = 2$) places the following restrictions based on set $\Theta_{I,stat}^{\{1,3\}}$:

$$\begin{aligned} (9) &: P(y_0 = 1, y_1 = 0) + P(y_2 = 0, y_3 = 1) \geq 1 \qquad \gamma \leq 0 \\ (10) &: P(y_0 = 0, y_1 = 1) + P(y_2 = 1, y_3 = 0) \geq 1 \qquad \gamma \leq 0 \\ (11) &: P(y_0 = 1, y_1 = 1) + P(y_2 = 0, y_3 = 0) \geq 1 \qquad \gamma \geq 0 \\ (12) &: P(y_0 = 0, y_1 = 0) + P(y_2 = 1, y_3 = 1) \geq 1 \qquad \gamma \geq 0 \end{aligned} \tag{4.4}$$

and based on set $\Theta_{I,stat}^{\{2,3\}}$:

$$\begin{aligned} (9) &: P(y_1 = 1, y_2 = 0) + P(y_2 = 0, y_3 = 1) \geq 1 \qquad \gamma \leq 0 \\ (10) &: P(y_1 = 0, y_2 = 1) + P(y_2 = 1, y_3 = 0) \geq 1 \qquad \gamma \leq 0 \end{aligned} \tag{4.5}$$

All the identified sets in the paper use conditional choice probabilities of the form (as an example)

$$
\begin{aligned}
P(y_1 = 0, y_2 = 1|y_0, x) &\equiv p_2(0, 1|y_0, x) \\
P(y_1 = 1, y_2 = 0|y_0, x) &\equiv p_2(1, 0|y_0, x) \\
P(y_1 = 0, y_2 = 0|y_0, x) &\equiv p_2(0, 0|y_0, x)
\end{aligned}
$$

The idea the inference section is to first construct a confidence region for the choice probabilities above. Then, heuristically, a confidence region for the identified set can be constructed by using draws from the (standard) confidence region for the choice probabilities. The mechanics of this exercise exploits linear programs to check whether a particular parameter vector θ belongs to the identified set. We describe this procedure in more details next.

5.1 A Confidence Region for the Choice Probabilities

One way to construct a confidence region for $p(y_0, x) = (p_2(0, 0|y_0, x), p_2(0, 1|y_0, x), p_2(1, 0|y_0, x))$ is as follows. Let $(y_0^1, x^1), \ldots, (y_0^J, x^J)$ denote the support of (y_0, x). Then, as sample size increases, we have

$$
\sqrt{n}W(p(\cdot)) \equiv \sqrt{n} \left(\begin{array}{l} (\frac{1}{n} \sum_i \hat{w}_i^{1,0}(y_0^1, x^1) - p_2(1, 0|y_0^1, x^1))1\{0 < p_2(1, \end{array} \right.
$$

and

$$\hat{p}_z(y_0, x) = \frac{1}{n} \sum_i 1\{y_{0i} = y_0, x_i$$

5.2.1 Linear Program for solving Model (5.2):

Conditions (5.2) are straightforward to verify for a given $p(\cdot)$ $CS^{p}_{1-\alpha}$. The following is an

$M_n(x,$

especially if the vector β is of high dimensions.

5.2.2 Linear Program for solving Model (5.3)

Now, building a CS for θ in the exchangeable model of (5.3) is more complicated since checking that both $-\Delta x \beta + \gamma y$

$1\{P_{(k)}(y_1 = 1, y$

income, $hinc_{it}$, is in dollar per month and is positive for all i and t. There are also time-inference section above and obtain a confidence region, we require that one obtains draws from the confidence region of the choice probabilities in (5.1) above. One computationally automatic way to get such draws is to use the Bayesian bootstrap which is equivalent to drawing from the posterior distribution of a multinomial with the usual Dirichlet priors[9]. For each draw from this posterior, we solve the linear program for min / max of the scalar a

mainly tied to establishing the empirical content of varying assumptions in dynamic binary choice models.

In establishing our theoretical results we reached the following conclusions regarding identifying the structural parameters in the model:

- Regression coefficients on strictly exogenous variables were generally easier to identify than the coefficient on the lagged binary dependent variable, which was our measure of the persistence in the model.
- Restricting dependence structure on the idiosyncratic components of the model facilitated identification of the structural parameters.
- Increasing the richness of the support of the exogenous variables facilitates the identification.
-

We demonstrate identification graphically with projections of three dimensional plots of our objective function. Specifically we look at values of the objective function of different values of β and γ along a grid of a two dimensional plane. Instead of constructing three dimensional plots, we show values of the parameters which attain the global maximum of the objective function. The objective functions used corresponded to the moment inequalities used in the main theorems. In models where point identification is attainable, a single value will be in the plot, whereas in partially identified models, a subset of the grid will be plotted.

6.2.1 Stationary Model, T=2

In this model we simulated data where v_{it}, x_{it} were each discretely distributed, with the number of support points for v_{it}, increasing from 2 to 7, and then continuously (standard normal) distributed. The number of support points for x_{it} was always two, though there where two distinct designs- one with identical support in each time period, and the other with strictly nonoverlapping support- $x_{it} = t \quad t = 1, 2$. The idiosyncratic terms u_{it} were bivariate normal, mean 0 variance 1, correlation 0.5, and the fixed effect α_i was standard normal. We assumed that all variables were mutually independent. The parameters where set to 1 for β and either 0.5 or -0.5 for γ.

Our plots for this model agree with our theoretical results. We note that when x_{it}, v_{it} are discrete, neither parameter is point identified. For example, in Figure 2, we have x is binary while v starts out as binary and then we add points of support ending with 14. This Figure is repeated for when true γ is negative. As we can see the identified set is not trivial. Its size shrinks in Figure 4 when v is normally distributed with increasing variance. Notice here that in all the plots, β appears well identified.

In Figure 6, we change x to a time trend ($x = t$) and in the top lhs plot, we have the identified set in the case when v is binary. Here, we cannot pin down the sign of γ. But, as we increase the points of support for v, the identified set shrinks and eventually it appears that the sign of γ is identified. The same story holds for when γ is negative. The next Figures allow for time trend in the case when v is normal.

Throughout, when v_{it} is continuously distributed, β is point identified, whereas γ is not. But the graph clearly demonstrates that its sign is.

the case when there was strictly nonoverlapping support conditions on x_{it}. In particular, Figure 17 shows that the identified set is essentially a point when v is normal, but y is not point identified when both v and x are discrete.

7 Conclusion

This paper analyzes the identification of slope parameters in panel binary response models with lagged dependent variables under minimal assumptions. In particular, we consider stationarity and exchangeability and characterize the identified set under these two restrictions without making any assumptions on the fixed effect. We show that the characterization yields the sharp set. In addition, we provide sufficient conditions for point identification even in models that have time trends as regressors, which is ruled out in Honoré and Kyriazidou (2000). The analysis is interesting and highlights the interplay between the strength of the assumptions, the number of time periods and the support of the exogenous regressors. Overall, we generalize many existing results for this model in interesting directions.

References

Aguirregabiria, V., J. Gu, and Y. Luo (2018): "Su□cient Statistics for Unobserved Heterogeneity in Structural Dynamic Logit Models," working paper, University of Toronto.

Altonji, J., and R. Matzkin (2005): "Cross Section and Panel Data Estimators for Nonseparable Models with Endogenous Regressors," *Econometrica*, 73, 1053–1102.

Arellano, M., and

Honoré, B., and E. Kyriazidou (2000): "Panel Data Discrete Choice Models with Lagged Dependent Variables," *Econometrica*, 68, 839–874.

——— (2019): "Identification In Binary Response Panel Data Models: Is Point-Identification More Common Than We Thought?," working paper.

Honoré, B., and A. Lewbel (2002): "Semiparametric Binary Choice Panel Data Models without Strictly Exogeneous Regressors," *Econometrica*, 70(5), 2053–2063.

Honoré, B., and E. Tamer (2006): "Bounds on Parameters in Panel Dynamic Discrete

Ouyang, F., S. Khan, and E. Tamer (2017): "Identification and Estimation of Dynamic Panel Data Multinomial Response Models," working paper, Duke University.

Pakes, A., and J. Porter (2014): "Moment Inequalties for Multinomial Choice with Fixed Effects," Harvard University Working Paper.

Polyakova, M. (2016): "Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D," *American Economic Journal: Applied Economics*, 8(3), 165–95.

Raval, D., and T. Rosenbaum (2018): "Why Do Previous Choices Matter for Hospital Demand? Decomposing Switching Costs from Unobserved Preferences," *Review of Economics and Statistics*, forthcoming.

Shi, X., M. Shum, and W. Song (2018): "Estimating Semi-Parametric Panel Multinomial Choice Models Using Cyclic Monotonicity," *Econometrica*, 86(2), 737–761.

Torgovitsky, A. (2016): "Nonparametric inference on state dependence with applications to employment dynamics," working paper, University of Chicago.

Yang, J., Y. Qi, and R. Wang (2009): "A Class of Multivariate Copulas with Bivariate Fréchet Marginal Copulas," *Insurance: Mathematics and Economics*, pp. 139–147.

A Figures

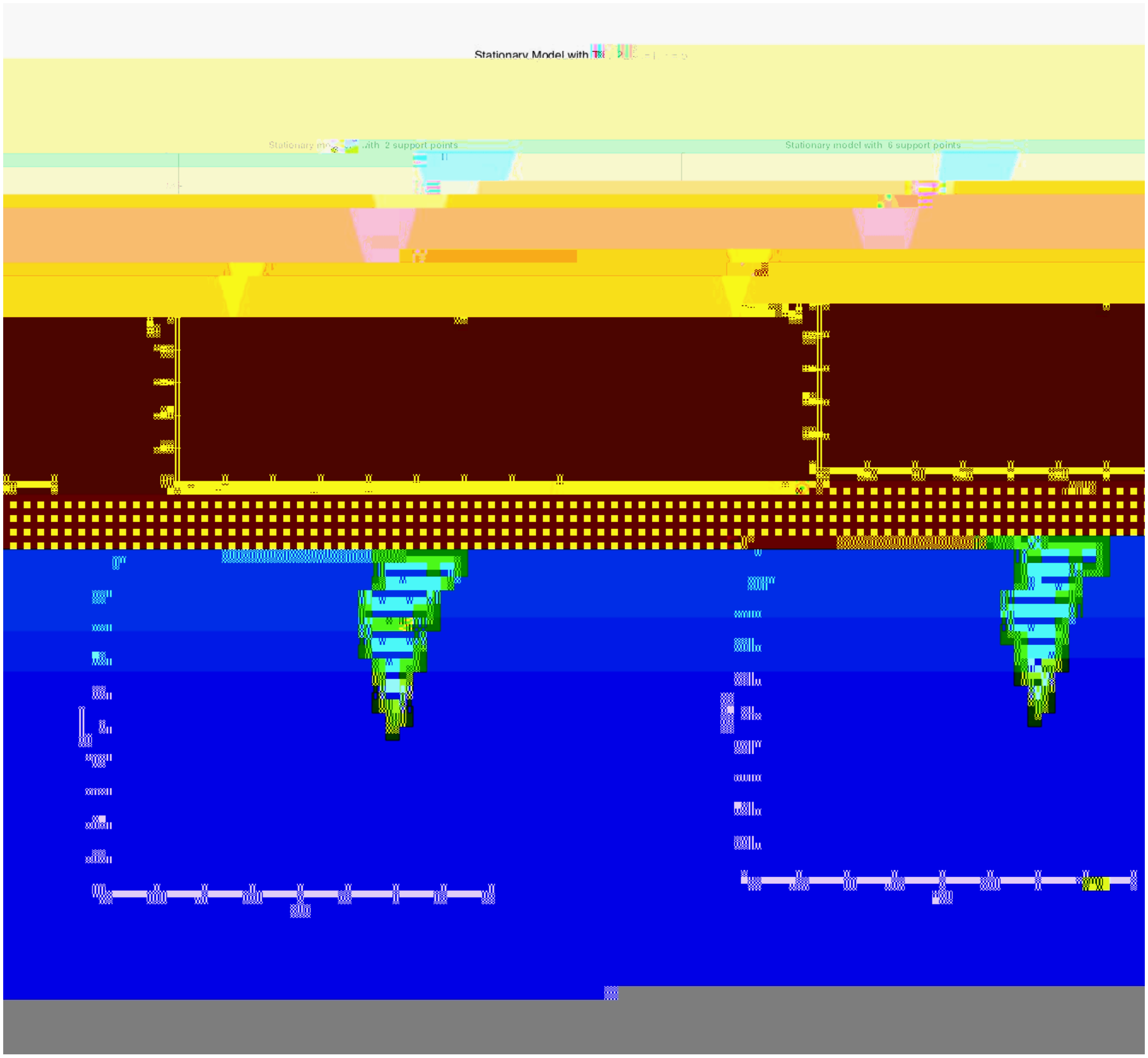


Figure 2: Stationary with $T = 2$ and Discrete Support with $\gamma = .5$

Figure 3: Stationary with $T = 2$ and Discrete Support with $\gamma = .5$

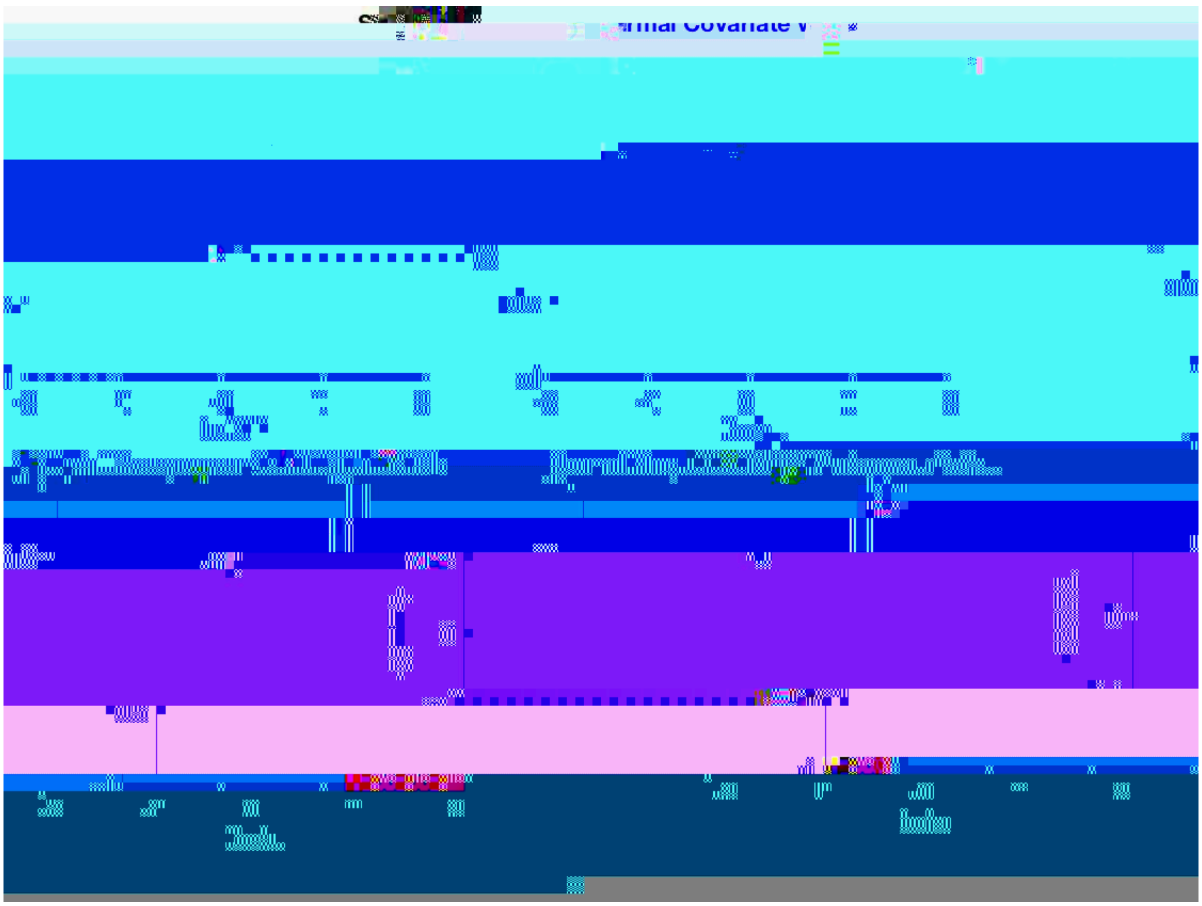


Figure 4:

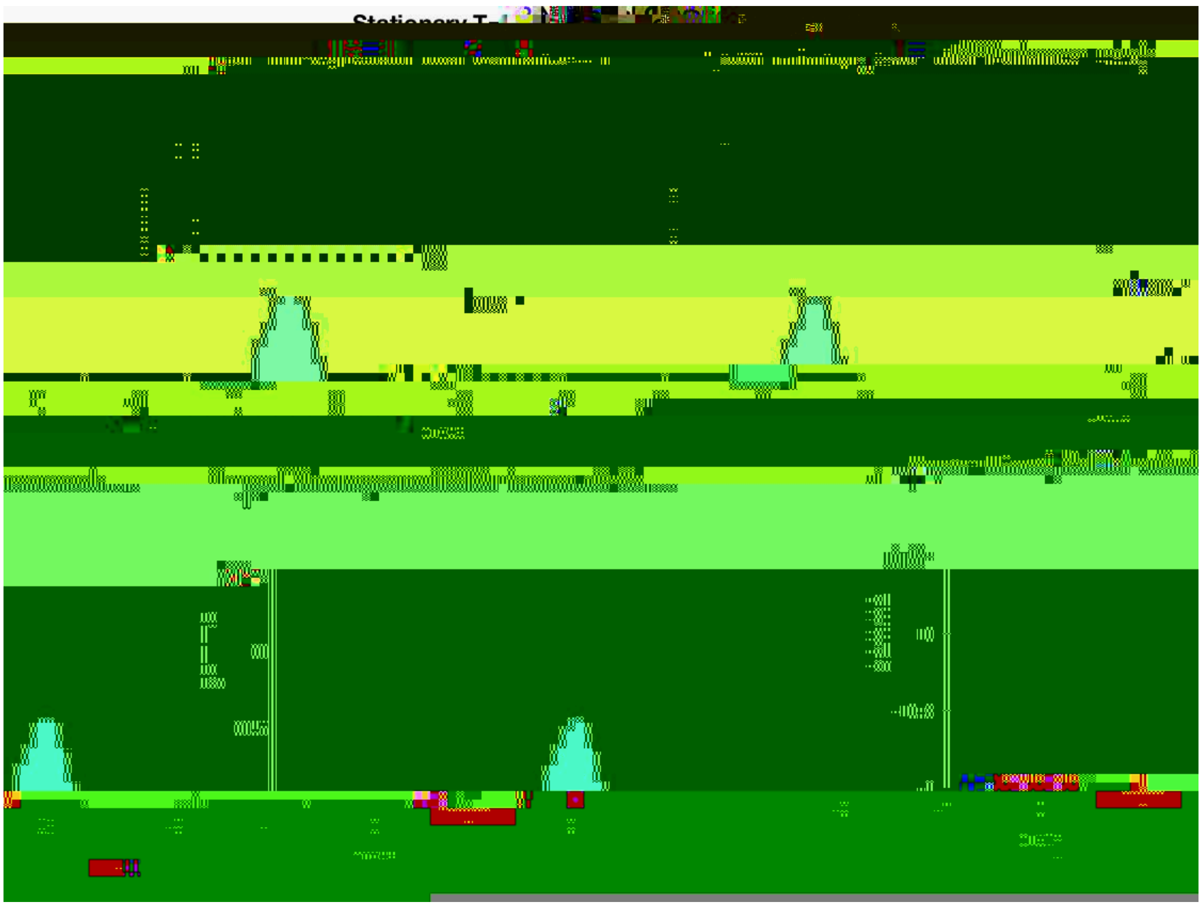


Figure 5: Stationary with $T = 2$ and Normal v with $\gamma = -.5$

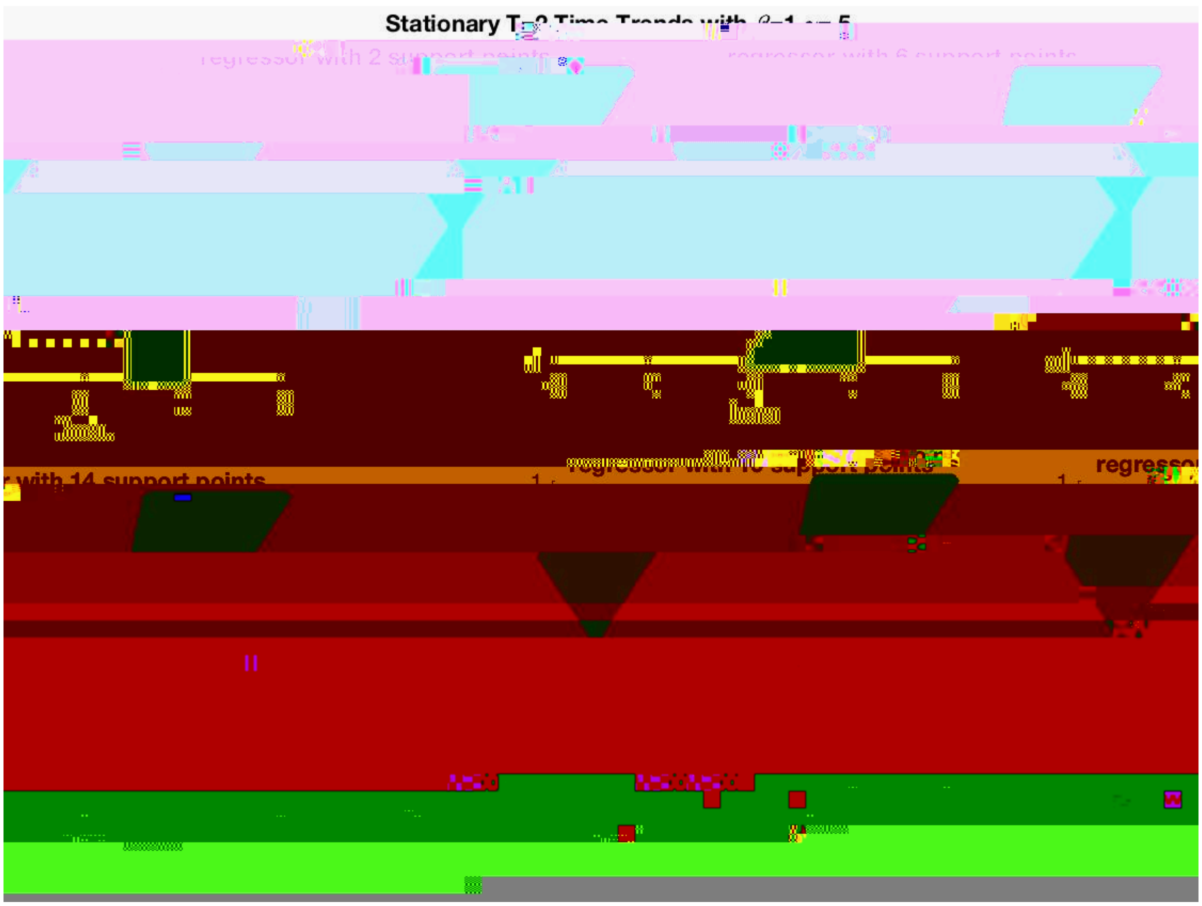


Figure 6: Stationary with $T = 2$ and Time Trend and Discrete Support for v with $\gamma = .5$

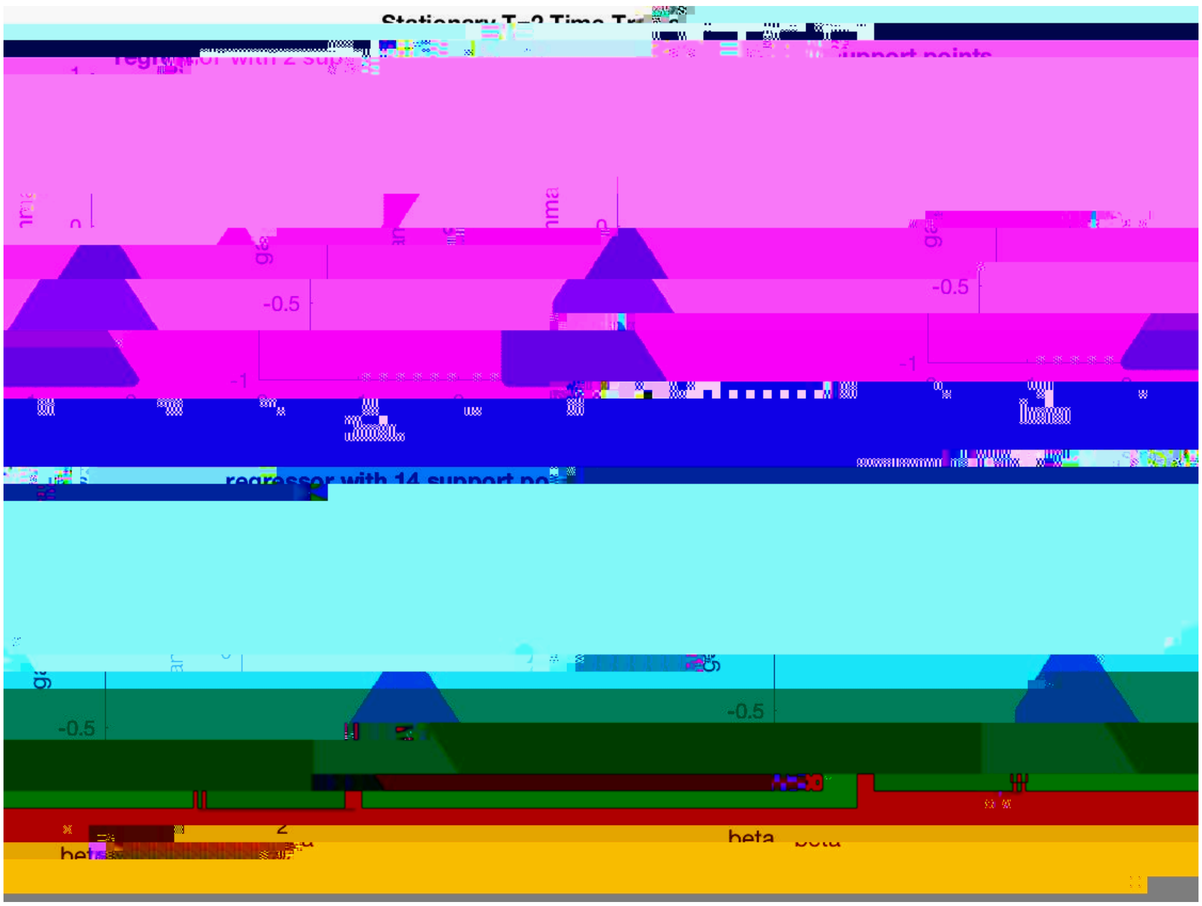


Figure 7: Stationary with $T = 2$ and Time Trend and Discrete Support for v with $\gamma = -.5$

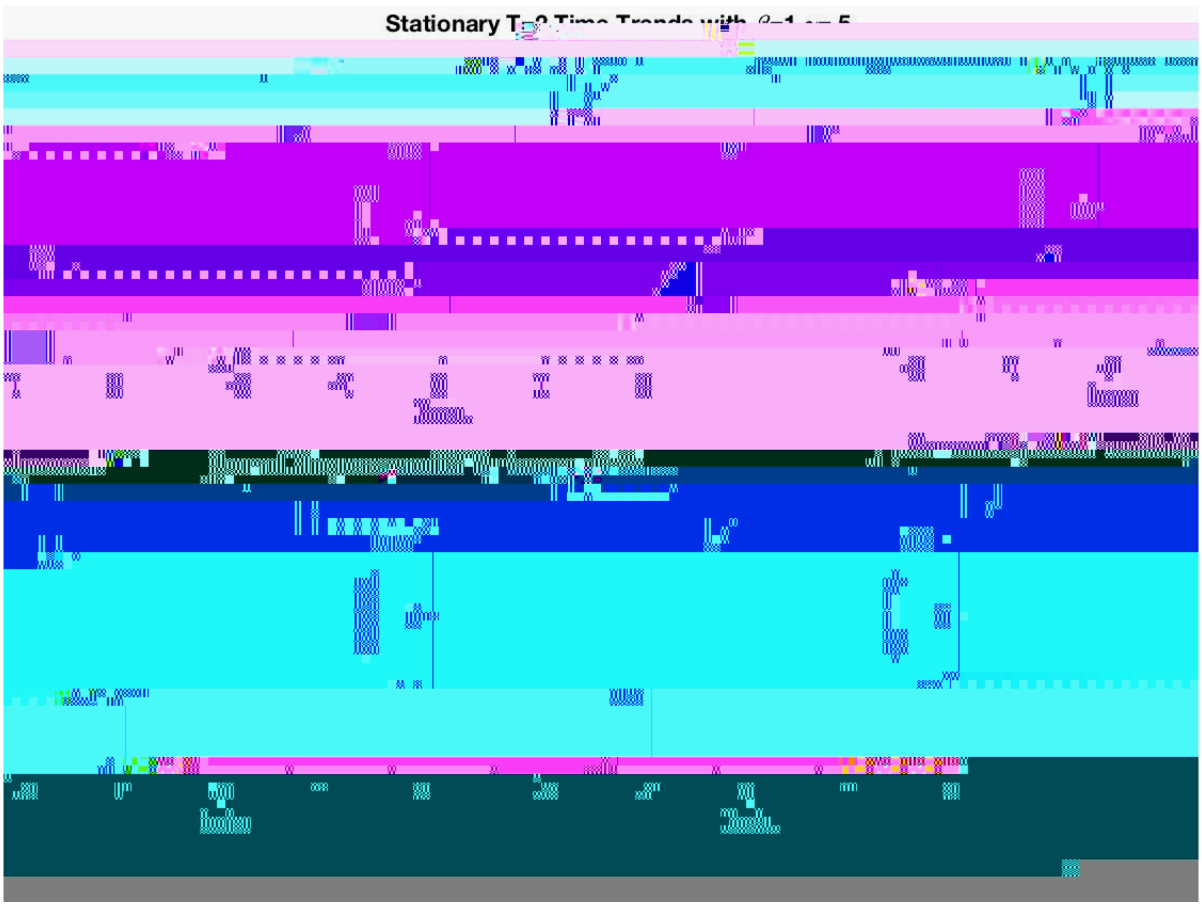


Figure 8: Stationary with T = 2 and Time Trend and Normal v with y = .5

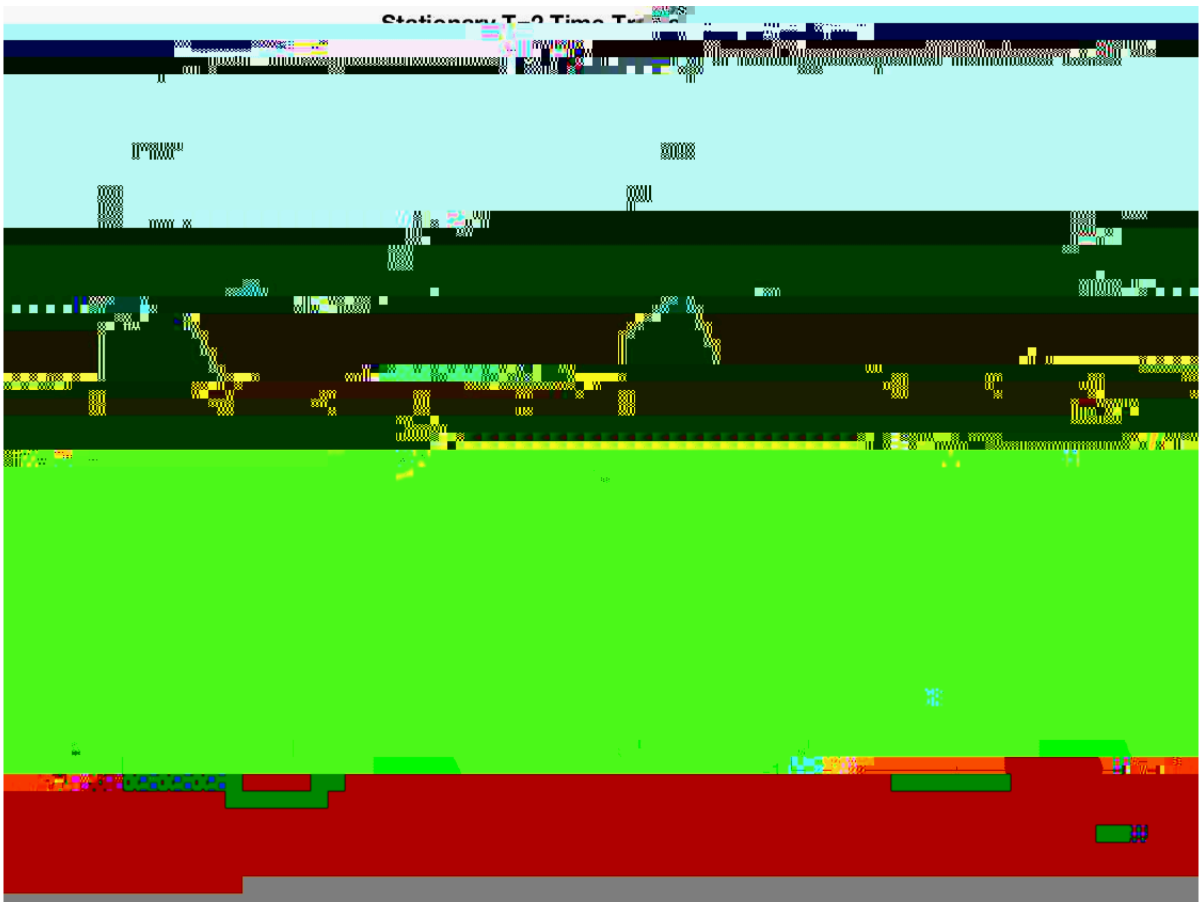


Figure 9: Stationary with $T = 2$ and Time Trend and Normal v with $\gamma = -.5$

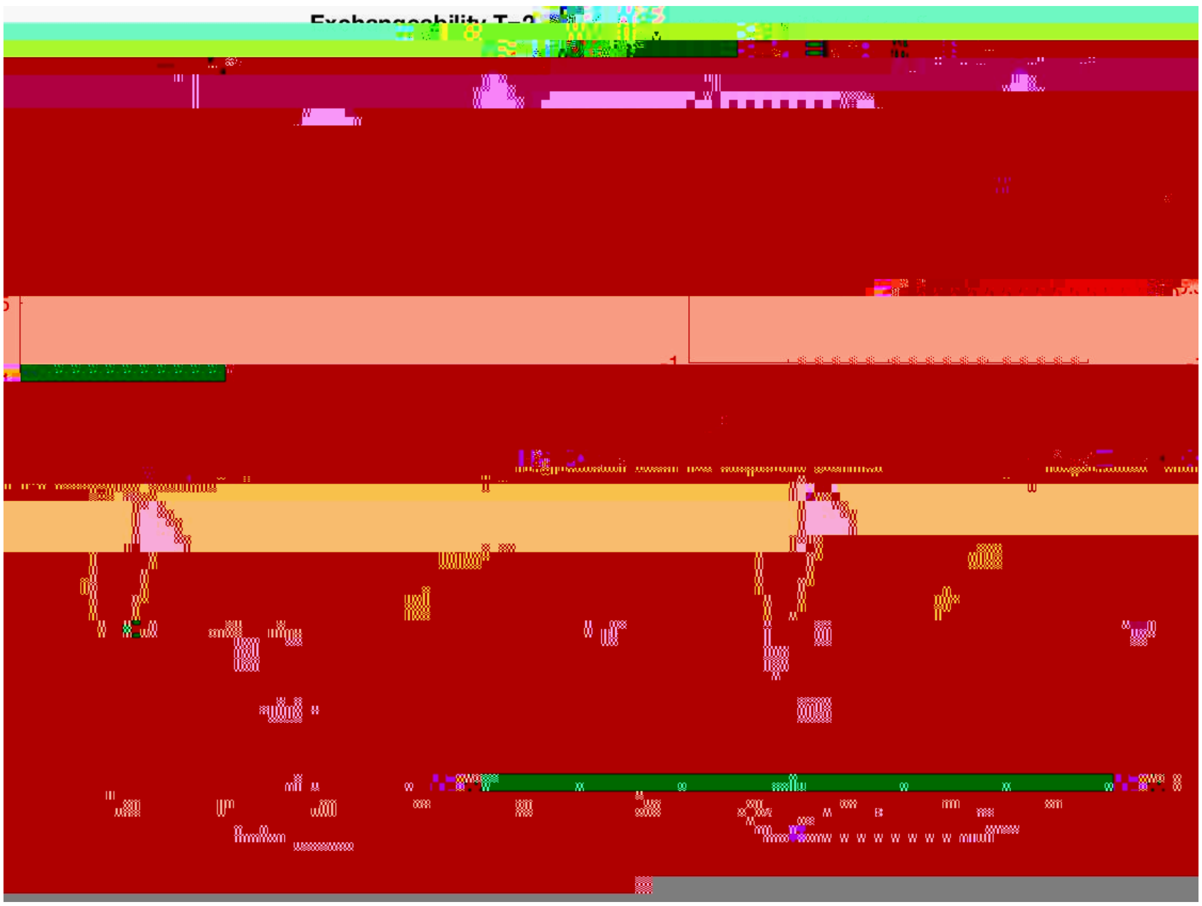


Figure 10: Exchangeability with $T = 2$ Discrete Support for v with $\gamma = .5$

Figure 11: Stationary with $T = 2$ Discrete Support for v with $\gamma = -.5$

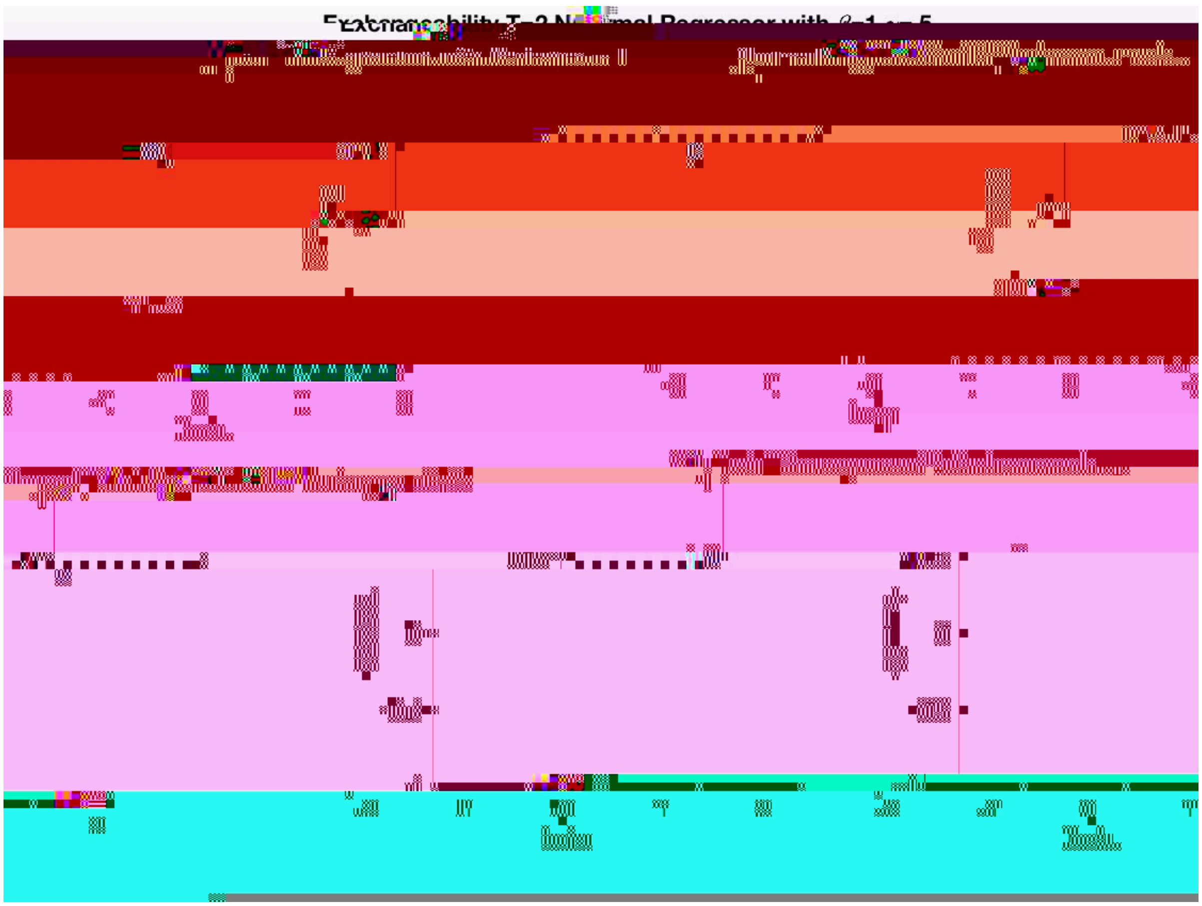


Figure 12: Exchangeability with $T = 2$ Discrete v with $\gamma = .5$

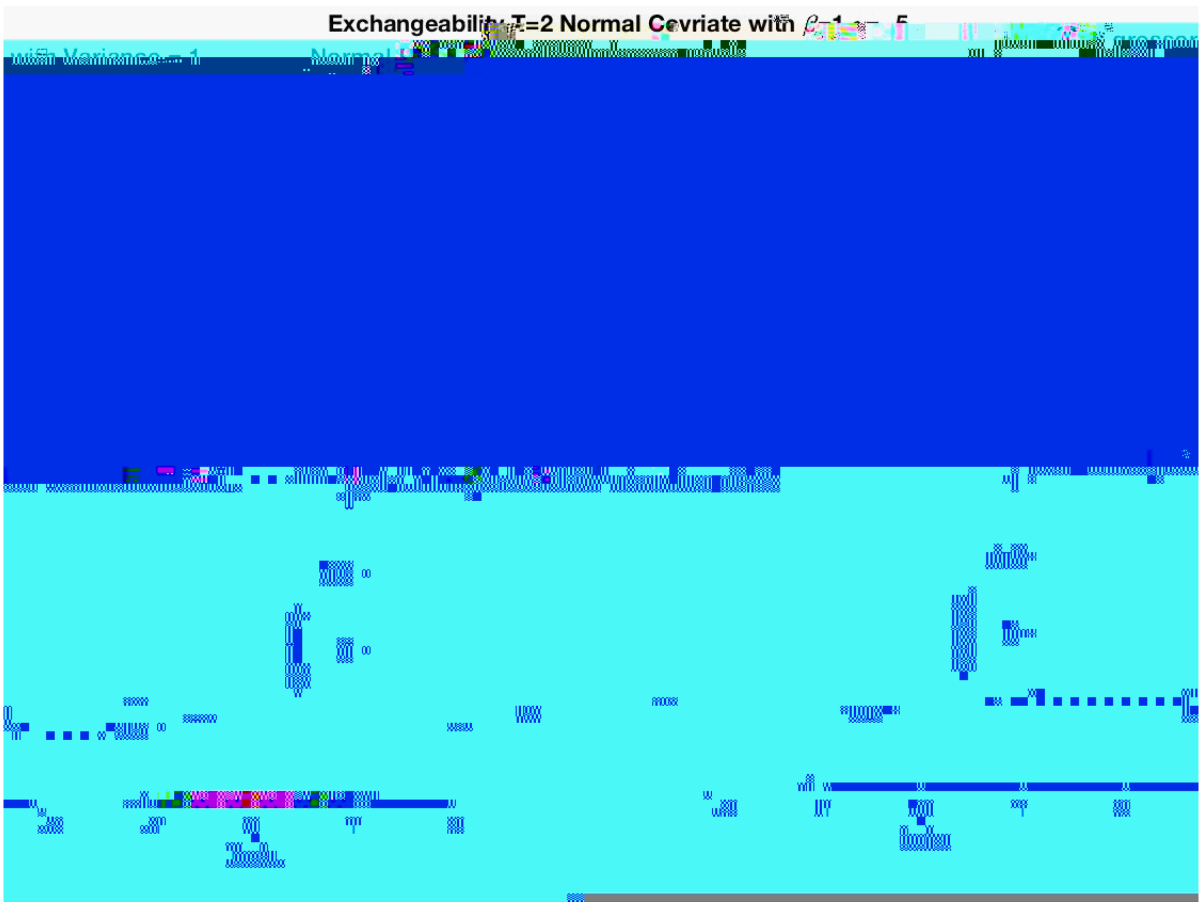


Figure 13: Stationary with $T = 2$ Normal for v with $y = -.5$

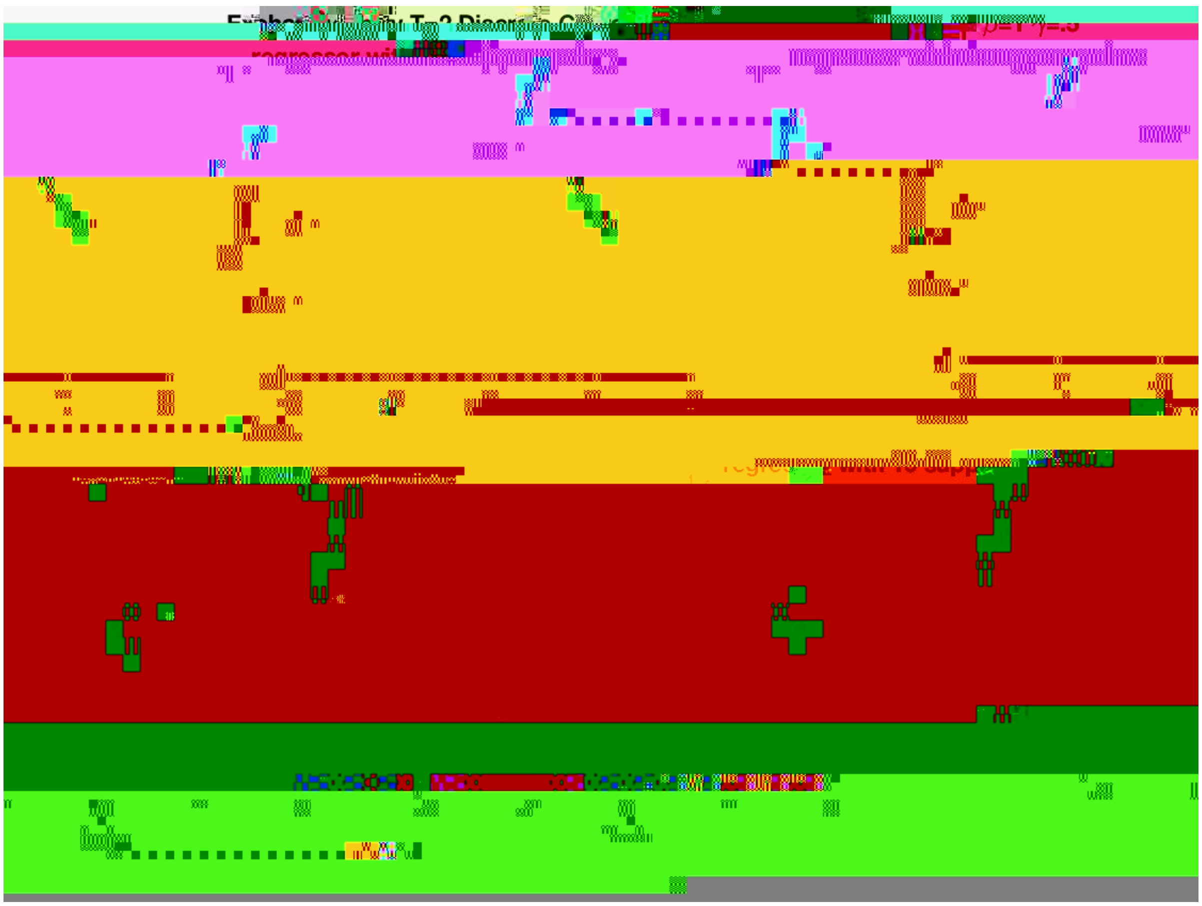


Figure 14: Exchangeability with $T = 2$ $x = t$ Discrete v with $\gamma = .5$

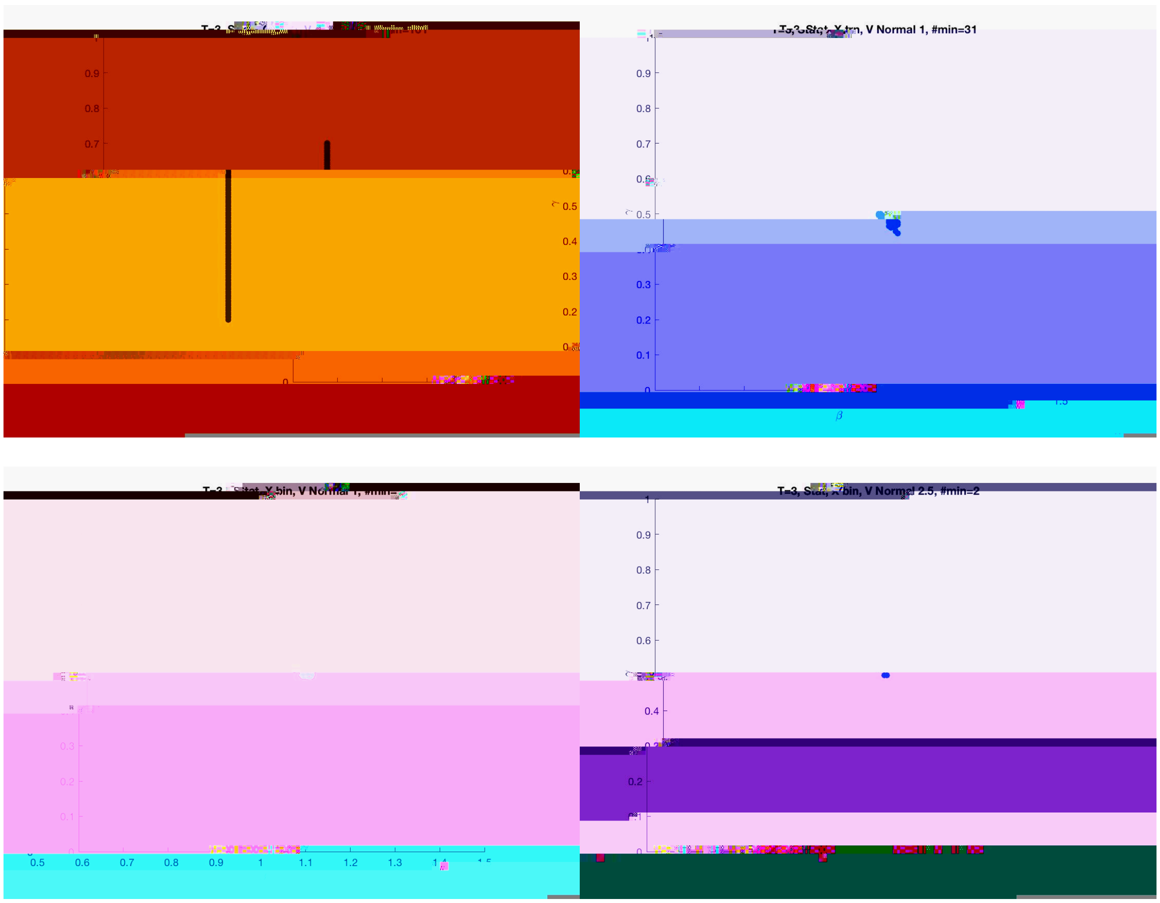
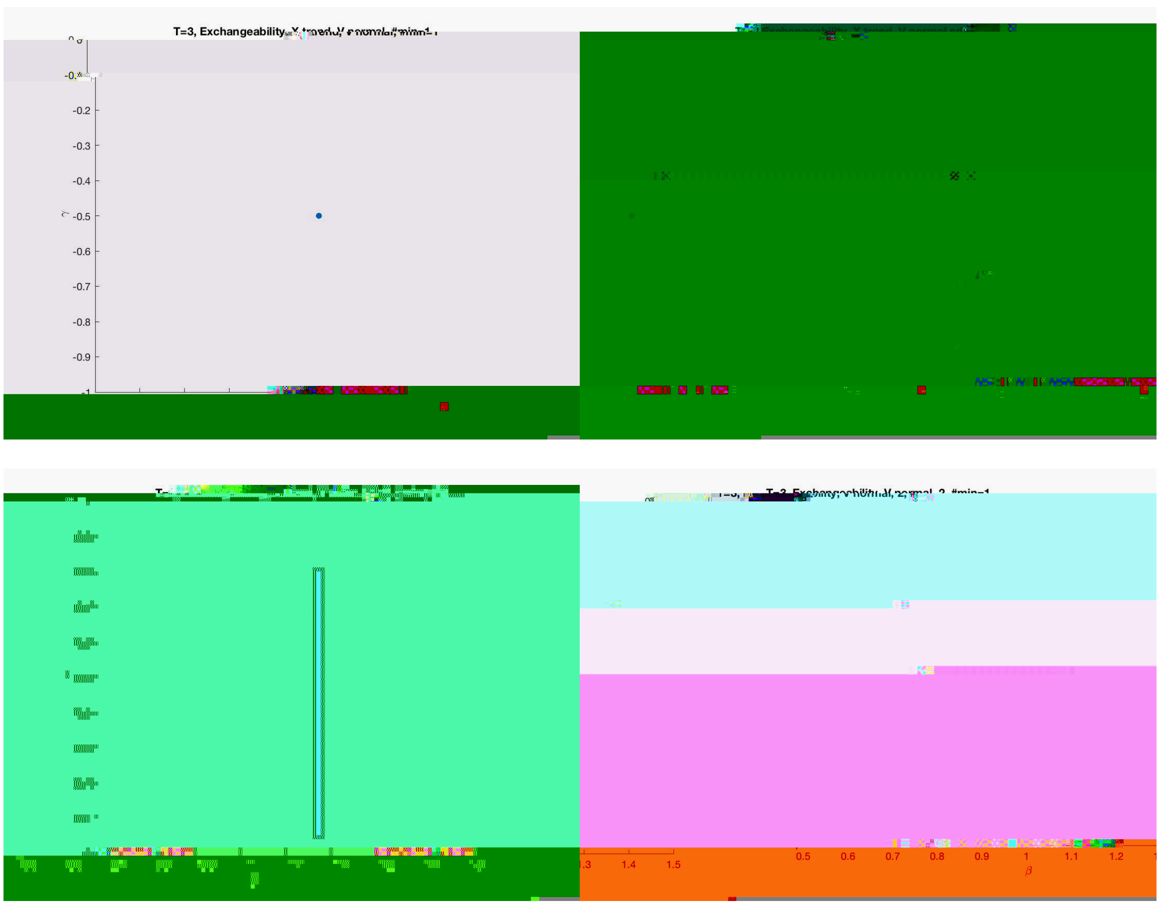


Figure 16: Stationarity with T = 3: Various Designs



T=3, Exchangeability, X trend, Y normal, #min=1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
-0.9
-1

B Proof

B.1 Proof of Lemma 3.1

Since $(u_{-M}, \ldots, u_0, u_1, \ldots, u_T)$ is exchangeable conditional on $\alpha, x_{-M}, \ldots, x_0, x_1, \ldots, x_T$, The-be the marginal distribution of v_t for $t = 1, 2$, conditional on x. The following are sharp and $\tilde{F}(\cdot|x)$ is the conditional distribution of some $\tilde{v}_t$. Let $\tilde{\alpha}$ be some scalar random variable (or even a constant), and define ˜

To sum up, for all x, y_0 in the support we have the following:

$$P(y_1 = 1, y_2 = 0|x, y_0) \leq F_{12}((x_1 - x_2)\beta + \gamma(y_0 - 1))$$
$$F_{12}((x_1 - x_2)\beta + \gamma y_0) \leq 1P(y_1 = 0, y_2$$

when the sign of γ can be identified. First note that if $\gamma \geq 0$, then Theorem 3.1 implies that

$$\Delta$$

B.4 Proof of Theorem 3.4

restrictions for v_2. In particular, $\Theta_{I,stat}^{\{2,3\}}$ is given by the restrictions: if for some x,

(1) $P(y_3 = 1|x) \geq P(y_2 = 1|x)$ $\quad (x_3 - x_2)\beta + |\gamma| \geq 0$;

(2) $P(y_2 = 1|x) \geq P(y_3 = 1|x)$ $\quad (x_3 - x_2)\beta - |\gamma| \leq 0$;

(3) $P(y_2 = 0, y_3 = 1|x) \geq P(y_2 = 1|x)$ or $P(y_1 = 1, y_2 = 0|x) \geq P(y_3 = 0|x)$ $\quad (x$