Economics Letters 113 (2011) 172–175

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

A nonparametric test for path dependence in discrete panel data

Maximilian Kasy *

Department of Economics, University of California - Los Angeles, 8283 Bunche Hall, Mail Stop: 147703, Los Angeles, CA 90095, United States

ARTICLE INFO

Article history:
Received 2 April 2010
Received in revised form
2 July 2011
Accepted 14 July 2011
Available online 23 July 2011

JEL classification:
C12
C14
C23

Keywords:
Path dependence
Mixtures of Markov chains
Nonparametric testing

ABSTRACT

This paper proposes a test for path dependence in discrete panel data based on a characterization of stochastic processes that are mixtures of Markov chains. This test is applied to European Community Household Panel data on employment histories. The data allow to reject the null of no path dependence in all subsamples considered.

1. Introduction

Path dependence is of potential relevance in many areas of economics and the social sciences more generally. Path dependence here is understood to signify a causal impact of past states of some system on the future of that system, holding the present state constant. For instance, the employment history of an individual i might have a causal impact on that individual's chance of finding a job, given present unemployment. This is suggested by the empirical observation that past employment status $Y_{i,0}$ is predictive of future status $Y_{i,2}$, conditional on present status $Y_{i,1}$, in panel data on individual employment histories. However, if there is unobserved and exogenous heterogeneity across individuals that is serially dependent, and influences employment prospects, a similar implication for observable data follows.

Several different approaches can be taken to identify the nature of path dependence in the presence of unobserved heterogeneity. Experimental variation of initial $Y_{i,0}$ identifies path dependence as the excess causal impact of $Y_{i,0}$ on $Y_{i,2}$, beyond the effect mediated through $Y_{i,1}$. The latter is identified by compounding the effect of $Y_{i,0}$ on $Y_{i,1}$ and the effect of $Y_{i,1}$ on $Y_{i,2}$.

Functional form assumptions underlie popular models of panel data as well as duration data. For instance, additive separability of heterogeneity is required in fixed effects models (see Chamberlain, 1985), and multiplicative separability of heterogeneity is imposed in the mixed proportional hazards model (see Heckman and Singer, 1985 and Van den Berg, 2001).

Without either experimental variation or functional form restrictions, models with arbitrary unobserved heterogeneity but no path dependence are still testable. In the case of spell durations, Heckman et al. (1990) devise tests based on characterizations of mixtures of exponential distributions. In the case of discrete panel data, Lee (1987) discusses restrictions on the coefficients of log–linear probability models implied by mixture assumptions. The present paper is based on a characterization of mixtures of Markov chains, proven by Diaconis and Freedman (1980). This characterization implies that, under the null of no path dependence, certain sequences of states have to occur with equal probability. The test proposed is a modified χ^2 test of these equality restrictions on the distribution of state sequences.

2. The test for path dependence

The time path $(Y_{i,t})$ of an individual's status is described by a Markov chain, conditional on time invariant individual specific heterogeneity α_i, if two assumptions hold. First, the conditional distribution of future status given the individual's history and time invariant exogenous characteristics does not depend on the individual's history: $P(Y_{i,t+1}|\alpha_i, Y_{i,t}, Y_{i,t-1}, \ldots) = P(Y_{i,t+1}|\alpha_i, Y_{i,t})$. This is implied by the absence of both path dependence and time varying heterogeneity. Second, this conditional distribution does not depend on time t. This paper proposes a test for the hypothesis that individuals' histories follow a Markov chain, conditional

* Tel.: +1 510 717 7012.
E-mail address: maxkasy@econ.ucla.edu.

0165-1765/$ – see front matter
doi:10.1016/j.econlet.2011.07.005

on time invariant individual specific heterogeneity. This implies that the population distribution of histories can be represented as a mixture of Markov chains.

Throughout, we consider discrete panel data with finite support, $Y_{i,t} \in \{y^1, \dots, y^m\}$. The event $\{Y_t = \sigma_t : t = 0, \dots, T\}$ is denoted by A_σ. The null hypothesis for which a test statistic is developed is the hypothesis that the data are generated from a mixture of Markov chains. A process (Y_t) is called a *mixture of Markov chains* if its law can be represented by

$$P(A_\sigma) = \int_{\mathscr{P}} \prod_{t=0}^{T-1} p(\sigma_t, \sigma_{t+1}) \mu(\mathrm{d}p, \sigma_0) \tag{1}$$

for some mixing distribution μ on the set of stochastic matrices $\mathscr{P}$ and initial states y_0. In this definition, $p(\sigma_t, \sigma_{t+1}) = P(Y_{i,t+1} = \sigma_{t+1} | \alpha_i, Y_{i,t} = \sigma_t)$. Conditional on α_i (that is, p), the probability of a given sequence $(y_0, \dots, y_T)$ is the product of the probabilities of transitions from y_t to y_{t+1}, where these transition probabilities are statistically independent and constant over time. Given the initial state, the probability of such a sequence only depends on the number of transitions between any pair of states. It does not depend on the order of these transitions. Two sequences with the same initial state and number of transitions have the same probability. This equality is preserved under mixing. The test proposed below is a test for equality of these probabilities.

Formally, two finite sequences of states, σ and τ, are called equivalent if they start with the same state and they have the same number of transition counts from p to q for every pair of states p and q, that is they contain the ordered tuple pq the same number of times. A process is called *partially exchangeable*, iff for all equivalent strings σ and τ, $P(A_\sigma) = P(A_\tau)$. Consider the equivalent sequences 1011 and 1101. Partial exchangeability implies these two sequences to be equally likely. If there was negative duration dependence in state 1, the second sequence might be less likely.

By the above argument, any process that is a mixture of Markov chains is partially exchangeable. That the reverse also holds true was proven by Diaconis and Freedman (1980), in an extension of the classic de Finetti's theorem. Their equivalence result requires the additional assumption of recurrence, where a process (Y_t) is called *recurrent*, if it returns with probability one to its initial state. Diaconis and Freedman (1980) prove the following theorem.

Theorem 1. *Let (Y_t) be recurrent. Then it is partially exchangeable iff it is a mixture of Markov chains.*

For a balanced panel, testing partial exchangeability amounts to testing equality restrictions on the multinomial distribution of state *sequences* in the population.[1] This can be done, in principle, using a generalized likelihood ratio test for equality restrictions on a multinomial distribution:

$$X^2 := 2 \sum_\sigma N_\sigma \log \left(\frac{N_\sigma}{np_\sigma} \right) \xrightarrow{d} \chi^2_k \tag{2}$$

under partial exchangeability as $n \to \infty$, where n is the number of cross-sectional units, σ is an index ranging over all possible sequences, N_σ is the number of observations of type σ, p_σ are the maximum likelihood probabilities of sequences σ subject to the equality restrictions implied by partial exchangeability, and k are the number of linearly independent restrictions. The null can be rejected for large test statistics.

[1] The assumption of recurrence in Theorem 1 is only needed for the implication from partial exchangeability to representability as a mixture of Markov chains. Hence, rejection of partial exchangeability implies rejection of a mixture of Markov chains even without recurrence.

Table 1
Number of linearly independent restrictions implied by partial exchangeability and number of different possible sequences (in brackets).

Periods	States				
	2	3	4	5	6
3	0 (8)	0 (27)	0 (64)	0 (125)	0 (216)
4	2 (16)	6 (81)	12 (256)	20 (625)	49 (1296)
5	10 (32)	57 (243)	168 (1024)	370 (3125)	2925 (7776)
6	32 (64)	315 (729)	1368 (4096)	4742 (15 625)	29 720 (46 656)
7	84 (128)	1347 (2187)	8492 (16 384)	44 247 (78 125)	234 602 (279 936)
8	198 (256)	4983 (6561)	44 660 (65 536)	304 587 (390 625)	1 572 245 (1 679 616)
9	438 (512)	16 899 (19 683)	211 124 (262 144)	1 759 026 (1 953 125)	9 829 528 (10 077 696)
10	932 (1024)	54 387 (59 049)	932 168 (1 048 576)	9 353 166 (9 765 625)	59 932 614 (60 466 176)

Consider, for example, the case of $m = 2$ states and $T + 1 = 4$ periods. Under the null, the sequences $\sigma = 1011$ and 1101 have to occur with equal probability; the same holds for the sequences 0100 and 0010. Suppose that in a sample of $n = 200$ individuals, we observe these sequences with frequencies N_σ equal to 10–40. The restricted MLE for the first two sequences is then given by $p_\sigma = 15/200$ and for the other two by $p_\sigma = 35/200$. This implies that $X^2 = 2(10 \log(10/15) + 20 \log(20/15) + 30 \log(30/35) + 40 \log(40/35)) = 4.83$. Since the 95% critical value of the χ^2_2 distribution is given by 5.99, we would not be able to reject the null of no path dependence in this example.

The number of restrictions k implied by partial exchangeability is shown in Table 1. As can be seen, greater length of the panel increases the ratio of restrictions to possible sequences, k/m^{T+1}. Many states (large m) might be problematic since the number of possible sequences, m^{T+1}, explodes, thus making the probability of observing any particular equivalence class low.

The asymptotic χ^2 approximation might fail in practice for two related reasons. First, some equivalence classes have actual probability 0. Second, some equivalence classes have very few observations. According to Van der Vaart (2000, Chapter 17), the χ^2 approximation under the null is "good" if there are, in expectation, at least 5 observations per possible sequence σ.

The following modification of the test is asymptotically valid and gives significant finite sample improvements. Count the number of observed sequences falling into each equivalence class, and discard all classes that contain less than "5 times the number of cells in the class" observations. Calculate the generalized likelihood ratio test statistic of the restrictions on this subsample. Reject the null if the statistic exceeds the critical value of a χ^2 distribution with degrees of freedom corresponding to the number of implied restrictions in the classes retained in the sample.

Consider again the previous example, but suppose now that the frequencies N_σ of the four relevant sequences are given by 10, 22, 3 and 6. This implies that the second equivalence class has 9 observations for 2 cells and hence is discarded. The modified test statistic is given by $X^2 = 2(10 \log(10/16) + 22 \log(22/16)) = 4.61$. The 95% critical value of the χ^2_1 distribution equals 3.84, and we could thus reject the null at a level of 5% in this example.

The asymptotic validity of this approach can be seen as follows. Condition on the distribution across equivalence classes. Calculate the generalized likelihood ratio test statistic for the null of a uniform distribution within each equivalence class. Note that

Table 2
Column (1) shows the number of observations in the subsample of each row, column (2) shows the number of remaining observations after discarding all sequences that lie in equivalence classes with too few observations, column (3) shows the fully unrestricted multinomial likelihood, column (4) shows the maximal likelihood under the restriction of partial exchangeability, column (5) shows the corresponding χ^2 statistic $\chi^2 = 2\sum_\sigma N_\sigma \log\left(\frac{N_\sigma}{np_\sigma}\right)$ where the p_σ are the restricted estimates, column (6) contains the number of equality restrictions implied by partial exchangeability in the subsample after discarding and column (7) contains the p-values from a χ^2 distribution corresponding to this number of restrictions.

χ^2 test for partial exchangeability

	(1) Observations	(2) Observations after discarding	(3) Unrestricted likelihood	(4) Restricted likelihood	(5) χ^2 statistic	(6) d*f* after discarding	(7) p-value
Germany	21,544	21,040	−60,696	−60,890	387.45	97	0.0000
Denmark	5,503	5,043	−15,511	−15,636	248.26	37	0.0000
The Netherlands	9,151	8,663	−27,148	−27,276	256.18	52	0.0000
Belgium	6,454	6,093	−17,834	−17,887	107.40	31	0.0000
Luxembourg	8,754	8,444	−23,253	−23,283	59.89	19	0.0000
France	13,306	12,710	−41,317	−41,453	271.10	66	0.0000
United-Kingdom	17,211	16,731	−45,570	−45,602	65.20	48	0.0498
Ireland	8,531	8,127	−26,102	−26,137	71.34	29	0.0000
Italy	17,780	17,172	−56,928	−57,100	344.09	88	0.0000
Greece	12,271	11,741	−38,270	−38,522	504.23	73	0.0000
Spain	16,263	15,528	−55,144	−55,435	582.11	122	0.0000
Portugal	11,858	11,243	−33,905	−33,984	158.36	59	0.0000
Austria	7,437	7,006	−21,442	−21,545	207.19	35	0.0000
Male, all countries	75,081	74,132	−243,400	−244,180	1576.20	194	0.0000
Female, all countries	80,982	80,185	−273,530	−274,390	1727.90	225	0.0000

for each of these test statistics, standard χ^2 asymptotics apply, conditional on the distribution across classes. Note, finally, that these statistics are conditionally independent across equivalence classes, and that the sum of independent χ^2 variables is χ^2 itself.

3. Application to employment data

We shall now apply this test to panel data on employment histories in Western European countries. The data set used is the ECHP household panel for the years 1995–2001. Individuals are coded to be in one of three states, employed, unemployed or unobserved. Recall that we defined the test statistic for balanced panel data. Here, we apply it to unbalanced data by coding unobservability as a distinct state, which makes the data set balanced, and test for path dependence in the process determining employment *and* observability. Since in our data unobservability is an absorbing state, any rejection of the null will still be driven by the employment process.

After discarding all individuals where first period (1995) employment status is unobserved, we get 156 060 sequences of 7 periods (years) and 3 states. Discarding based on initial state does not affect the validity of the test, as its asymptotic approximation is valid *conditional* on initial states.[2]

The results of applying the modified χ^2 test for partial exchangeability to these data are shown in Table 2. In all cases but the UK we get p-values far below 1%, and even for the UK we are below 5%. Only a small fraction of the observations have to be discarded for the modified test. The null of a mixture of Markov chains can be rejected in all subsamples considered.

This application illustrates the weak data requirements of the test proposed. It can be applied to any discrete panel data set. The panel need not be balanced. Spells can be both left and right censored without affecting the validity of the test. No covariates, whether time invariant or changing over time, are needed.

4. Discussion and conclusion

We have tested and rejected the null that individual employment histories are generated from a mixture of Markov chains. This null, and the notion of path dependence more generally, are relative to the coding of status. If the causal effect of $Y_{i,0}$ on $Y_{i,2}$ is mediated through $Y_{i,1}$ and $Z_{i,1}$, including Z in the coding of status eliminates path dependence. In particular, even if data are generated from a Markov process, aggregation of states leads to violation of the Markovian property and hence of partial exchangeability.

The null also implies time homogeneity. This cannot be relaxed fully, since any distribution of sequences can be generated from a mixture of time inhomogeneous processes without path dependence. We could allow for aggregate structural breaks in an extension of the test. For instance, choose a breakpoint and calculate the previous test statistic for either part of the time window, take the sum and reject for the appropriate critical value of a χ^2 distribution with degrees of freedom equal to the sum of the number of restrictions from both parts. In another generalization, one can test for higher order Markovian behavior conditional on time invariant heterogeneity. By redefining states as $Z_t = (Y_t, Y_{t-1})$ for instance, second order Markov behavior of Y is equivalent to first order Markov behavior of Z. Applying either extension to the ECHP data again allows to reject the null hypotheses.

To conclude, it should be emphasized that the test proposed is a complement rather than a substitute for inference procedures relying on stronger assumptions. It is not able to disentangle the nature of path dependence (duration dependence in which state?), nor can we allow for time varying exogenous covariates, as d'Addio and Honoré (2006) do. It is attractive, however, because it requires neither functional form restrictions nor exogenous sources of variation.

Acknowledgments

This work was supported by a DOC fellowship of the Austrian Academy of Sciences at the Department of Economics, UC Berkeley as well as by the Jubilaeumsfonds of the Austrian Nationalbank grant no. 9632 which provided access to the data. I would like to thank Geert Ridder at USC, Helmut Hofer at the IHS Vienna and Bryan Graham, Michael Jansson, James Powell, Steven Raphael, Paul Ruud and Sarath Sanga at UC Berkeley for many valuable discussions.

[2] Note that we are *not* discarding observations with left-censored spells; potentially all initial spells can be left-censored.

References

Chamberlain, G., 1985. Heterogeneity, omitted variable bias, and duration dependence. In: Longitudinal Analysis of Labor Market Data. pp. 3–38.

d'Addio, A., Honoré, B., 2006. Duration Dependence and Timevarying Variables in Discrete Time Duration Models. Princeton University.

Diaconis, P., Freedman, D., 1980. de Finetti's theorem for Markov chains. The Annals of Probability 8 (1), 115–130.

Heckman, J., Robb, R., Walker, J., 1990. Testing the mixture of exponentials hypothesis and estimating the mixing distribution by the methods of moments. Journal of the American Statistical Association 85 (410), 582–589.

Heckman, J., Singer, B., 1985. Social science duration analysis. In: Longitudinal Analysis of Labor Market Data. pp. 39–110.

Lee, L., 1987. Non-parametric testing of discrete panel data models. Journal of Econometrics 34 (1–2), 147–177.

Van den Berg, G., 2001. Duration models: specification, identification and multiple durations. In: Handbook of Econometrics. pp. 3381–3460.

Van der Vaart, A., 2000. Asymptotic Statistics. Cambridge Univ. Press.