Modeling dynamic preferences. A Bayesian robust dynamic latent ordered probit model[1]

Forthcoming in *Political Analysis*

Daniel Stegmueller
Department of Government
University of Essex
Wivenhoe Park, Colchester CO4 3SQ, UK
mail@daniel-stegmueller.com

Abstract

Much politico-economic research on individuals' preferences is cross-sectional and does not model dynamic aspects of preference or attitude formation. I present a Bayesian dynamic panel model, which facilitates analysis of repeated preferences using individual-level panel data. My model deals with three problems. First, I explicitly include feedback from previous preferences taking into account that available survey measures of preferences are categorical. Second, I model individuals' initial conditions when entering the panel as resulting from observed and unobserved individual attributes. Third, I capture unobserved individual preference heterogeneity, both via standard parametric random effects, and via a robust alternative based on Bayesian nonparametric density estimation. I use this model to analyze the impact of income and wealth on preferences for government intervention using the British Household Panel Study from 1991–2007.

[1] I am indebted to Jeff Gill, Michael Malecki, Michael Becher, Tom Snijders, Simon Jackman, Thomas Gschwend, Vera Troeger, Adam Ziegfeld, Jeroen Vermunt, Martyn Plummer, as well conference and seminar participants in Chicago, Berlin, Cologne, Oxford, and Tilburg for helpful comments and criticisms. Equal thanks is due to my reviewers and the editors. Furthermore, I thank the Oxford Supercomputing Centre for resources and support.

1. INTRODUCTION

Individuals' political and economic preferences typically exhibit patterns of both stability and change (e.g. Wlezien 1995). On the one hand, preferences are often very highly correlated over time. But, on the other hand, preferences can change in response to external events, such as income shocks, becoming unemployed, or experiencing an economic crisis. To capture the dynamics of preferences – their stability and their change – an appropriate modeling strategy involves the use of individual-level panel data and dynamic panel models, in which past preferences influence current preferences via a first-order Markov process. Panel data are increasingly being used in political science, both in the form of long-term household panels, such as the British Household Panel Survey, and election panels, such as the Cooperative Campaign Analysis Project. *Linear* dynamic panel models are also well known in political science (for an introduction see Wawro 2002 in this journal). However, the application of these models to modeling dynamic preferences is not straightforward.

Three central issues arise when modeling preference dynamics: categorical preference measures, endogenous initial observations, and individual heterogeneity. First, although political scientists conceive of preferences as continuous, available survey data on preferences is usually ordered-categorical, often using rather coarse categories. The nonlinear nature of preference measures prohibits direct application of established linear dynamic panel models (e.g. Arellano and Bond 1991; Blundell and Bond 1998) and instead requires a dynamic model for categorical data for both the dependent variable and the feedback process. Second, because initial conditions – an individuals' preference states when entering the panel – are endogenous to the preference formation process under study, one should explicitly model initial conditions in nonlinear panel models (Heckman 1981b; Nerlove et al. 2008). Third, unobserved individual heterogeneity must also be modeled explicitly in order to capture unobserved or unmeasured effects of individual characteristics such as motivation or ability. When modeling heterogeneity via Gaussian random effects – as is standard in virtually all hierarchical models in political science – inferences can be sensitive to this specific distributional assumption and should be checked using a more flexible model specification.[2] Standard fixed effects estimation strategies are unavailable due to the presence of a lagged dependent (endogenous) variable in the nonlinear model (see, e.g. Nickell 1981; Heckman 1981b; Arellano and Carrasco 2003).

I present a Bayesian robust latent dynamic ordered probit model, which tackles these three problems. First, it captures the categorical nature of survey-based preference measures by using

[2]Dynamic panel models for ordinal data are not widely developed in political science. Theoretical work and applications exist in biostatistics, medicine, and finance (e.g. Lunn et al. 2001; Hasegawa 2009; Varin and Czado 2010; Czado et al. 2011; Müller and Czado 2005), but are developed with long time-series in mind, and are not concerned with initial conditions in short panels of individuals (note that the start of medical studies often does coincide with the start of the data generating process). Pang (2010) presents a model for repeated categorical data using correlated residuals. However, extending the model to include dynamic feedback is not straightforward due to the special status of initial conditions (cf. appendix A). Pudney (2006, 2008) presents a model for dynamic ordinal data using Gaussian random effects in a maximum likelihood framework.

an ordered probit specification, in which a continuous latent preference variable generates observed survey responses. Most existing categorical dynamic panel models specify the lagged dependent variable as categorical, which implies the unrealistic assumption that current continuous *preferences* are influenced by past categorical survey *responses*. In contrast, I specify feedback from previous preferences to current ones as also arising from latent preferences, thus appropriately distinguishing between continuous concept and categorical survey items. Second, I model initial conditions using a simultaneous equation specification, in which individuals' initial observations depend on observed covariates, background information, such as parents' education, and unobserved individual specific effects. Third, I present robust specifications for the distribution of unobserved heterogeneity. I specify hierarchical or multilevel models with both Gaussian and t-distributed random effects. To relax these parametric assumptions, I employ Bayesian nonparametric density estimation for flexible estimation of the random effects distribution using Dirichlet process priors (for recent applications of Bayesian nonparametrics in political science see Imai et al. 2008; Gill and Casella 2009; Grimmer 2010; Spirling and Quinn 2010).

The paper proceeds as follows. In the next section I set up the hierarchical latent dynamic panel model, discuss my treatment of initial conditions, the specification of priors, and possible model extensions. Next, I present robust random effects specifications using Dirichlet process priors. I illustrate the model by an example from the political economy of redistribution preferences – where studies are usually cross-sectional and ignore both unobserved heterogeneity and dynamics. I analyze the impact of income and wealth on preferences for government intervention using the British Household Panel Study from 1991–2007, which repeatedly measures individual preferences for nearly 2000 individuals. I discuss results arising from the model specification using standard Gaussian random effects and illustrate how to conduct robustness tests using the flexible Dirichlet process random effects model. The last section concludes the paper.

2. LATENT DYNAMIC MODEL

A dynamic analysis of individual behavior or preferences has three features not present in cross-sectional studies. First, individual preferences show a certain degree of persistence. While cross-sectional studies provide a snapshot of individuals in time, modeling the dynamics of preferences using panel data provides an explicit model of how preferences change over time (Bartels 1999). A straightforward theoretical specification posits that preferences are persistent, which creates correlated observations within the same individual. In other words, "[...] preferences remain unchanged unless something happens to change them [...]" (Wlezien 1995: 989). Thus a dynamic model of preferences should include a *persistence parameter* capturing this correlation.

Second, some individual characteristics, such as intelligence or motivation, can have a strong influence on preferences or attitudes, but are unobserved or unobservable to the researcher. This *individual heterogeneity* is captured via individual constants, which I specify as random effects (I discuss robustness of distributional assumptions in section 3). It is well known that if heterogeneity is present in the true data generating process but ignored in the estimated model, the degree of preference persistence will be overestimated (see Heckman 1981a). Conversely, ignoring persistence leads researchers to overstate the extent of heterogeneity. Thus, a completely specified model of dynamic preferences has to include both components.[3]

Third, a sample of individuals, be it cross-sectional or a panel, provides only a time-limited observation window. Individuals started forming their beliefs and preferences a long time before one starts observing them. The fact that individuals do not enter a study with an 'empty mind', i.e. the problem of initial conditions, has to be included in the model. Those three features are important, when interpreting the effect of shocks (such as becoming unemployed) on preferences. Estimating the effect of such shocks from cross-sectional data, ignoring preference persistence as well as individual heterogeneity, might lead to erroneous conclusions.

2.1. *Modeling dynamics*

Concepts like preferences and attitudes are not inherently discrete. The fact that one works with categorical variables is usually simply due to methodological limitations in data collection and measurement (McKelvey and Zavoina 1975). Consequently, preferences should be specified as a *latent variable* z_t which represents the underlying continuous concept that generates observed categorical scores y_t (e.g. Greene and Hensher 2010). Since from the conceptual perspective of preferences there is no reason to expect that current *continuous* preferences depend on past preference *categories*, we also need the latent variable to appear on the right hand side of our dynamic panel model (Heckman 1978; Müller and Czado 2005; Pudney 2008). In other words, feedback from past preferences to current ones, should be specified as arising from z_{t-1} not y_{t-1}.[4]

Thus, following Albert and Chib (1993), I model observed responses in category c ($c = 1, \ldots, C$) of observed variable y_{it} ($i = 1, \ldots, N; t = 0, \ldots, T$) as being generated by an underlying continuous latent variable z_{it} and a vector of threshold parameters $\boldsymbol{\tau}$ such that

$$y_{it} = c \text{ if } z_{it} \in (\tau_{c-1}, \tau_c]. \tag{1}$$

[3] The importance of distinguishing persistence (or state dependence) and heterogeneity has been well established in economics (e.g. Heckman 1981a; Keane 1997; Vella and Verbeek 1998; Arulampalam 2000). For recent discussions of its relevance to political science, see Wawro (2002) and Bartels et al. (2011).

[4] One of my reviewers rightly pointed out that other mechanisms could introduce dependence on past preferences, for example when individuals 'adapt' to repeatedly presented categories. If the objective of an analysis is to study these survey-method effects, the model can be extended, for example by including dummy response categories in addition to the latent variable (see Heckman 1978 for a detailed discussion of continuous and categorical lagged dependent variables).

To capture the ordinal nature of observed preference scores, threshold parameters are constrained to be monotonically increasing,

$$-\infty = \tau_0 < \tau_1 = 0 < \tau_2 < \cdots < \tau_{C-1} < \tau_C = \infty; \tag{2}$$

and $\tau_1 = 0$ to identify the model (assuming that an overall constant will be included in the model; see Albert and Chib 1993; Johnson and Albert 1999).

Now, the dynamic model for latent preferences z_{it} can be written as:

$$z_{it} = \phi z_{it-1} + \boldsymbol{\beta}' \boldsymbol{x}_{it} + \xi_i + \epsilon_{it}, \ t = 1, \ldots, T \tag{3}$$

where ϕ captures the degree of preference persistence, i.e. the extent to which current preferences depend on previous ones. $\boldsymbol{\beta}$ is a vector of regression parameters for matrix $\boldsymbol{x}_{it}$ of possibly time-varying covariates and an overall constant. Errors are decomposed into an individual-specific time constant random effect ξ_i and stochastic disturbances ϵ_{it}, which vary over individuals and survey waves. For identification, the variance of the stochastic errors, distributed $\epsilon_{it} \sim N(0, \sigma^2_\epsilon)$ has to be fixed. I set $\sigma^2_\epsilon = 1$, yielding an ordered probit specification.[5]

Unobserved individual heterogeneity is modeled via random effects, which are drawn from a normal distribution centered at zero with estimated variance σ^2_ξ:

$$\xi_i \sim N(0, \sigma^2_\xi). \tag{4}$$

The model can be seen as a multilevel or hierarchical model, with responses nested within individuals. The presence of random effects induces correlations between responses of the same individual over time (Rabe-Hesketh and Skrondal 2008).[6] The proportion of total variance that is due to individual random effects, after accounting for preference persistence, can be estimated by

$$\rho = \frac{\sigma^2_\xi}{(1 + \sigma^2_\xi)}. \tag{5}$$

This provides a useful indicator of the relevance of unobserved individual differences, ignored in cross sectional analyses.

[5] As usual, errors are assumed independent, $\text{Cov}(\epsilon_{is}, \epsilon_{it}) = 0 \forall s \neq t$, and uncorrelated with covariates, $\text{Cov}(\epsilon_{it}, x_{it}) = 0$.

[6] I employ standard assumptions of normal random effects, i.e. they are assumed to be independent of stochastic errors: $\text{Cov}(\xi_i, \epsilon_{it}) = 0$, and independent of $\boldsymbol{x}_{it}$: $\text{Cov}(\xi_i, \boldsymbol{x}_{it}) = 0$. The latter assumption is principally unverifiable. Thus Pudney (2008) suggests to regard this as a normalization and interpret effects of covariates $\boldsymbol{x}^*_i$ (those covariates in $\boldsymbol{x}_{it}$ which are time-constant) as combination of the true effect of $\boldsymbol{x}^*_i$ and the part of the random effect ξ_i that can be proxied by a linear function of $\boldsymbol{x}^*_i$. The estimated random effects variance σ^2_ξ is then interpreted as variation not predicted by $\boldsymbol{x}^*_i$. Alternatively, the model might be extended to allow for correlated random effects (Mundlak 1978; Wooldridge 2002).

2.2. *Modeling initial observations*

The previous discussion indicates that one generally assumes preference or attitude formation to be a continuous ongoing process. However, panel data provide only a limited window into this process. Clearly, the first panel observation of an individual does not coincide with the first time he or she has ever formed a preference. To the contrary, most researchers would argue that individuals start forming preferences at a very young age, and are influenced by parental characteristics, such as education, and by both observed and unobserved individual characteristics.[7] Thus, modeling initial observations has special relevance in a (short) dynamic panel model, as one's "assumption about the initial observations plays a crucial role in interpreting the model" (Anderson and Hsiao 1981: 598).[8]

Nerlove et al. (2008: 11-12) argue that initial observations should be modeled by a specification similar to the one affecting the remaining observations – i.e., as depending on observed individual characteristics in $\boldsymbol{x}_i$, while possibly including additional background variables $\boldsymbol{v}_i$, such as parental education or the region of upbringing. Furthermore, to capture the dependence of the initial observation on unobserved individual characteristics, one should specify an arbitrary correlation with the individual specific effect ξ_i (Nerlove et al. 2008; Harris et al. 2008). In specifying an explicit model for endogenous initial observations, I follow Heckman (1981a, b), who specifies an approximation for the fist (latent) observation $z_{i0}|\boldsymbol{x}_{it}, \xi_i$ as:

$$z_{i0} = \boldsymbol{\delta}'\boldsymbol{w}_i + \lambda\xi_i + \epsilon_{i0} \tag{6}$$

where $\boldsymbol{w}_i = (\boldsymbol{x}_{i0}, \boldsymbol{v}_i)$ is a vector of initial observation covariates comprised of an individual's covariate values at sample entry $\boldsymbol{x}_{i0}$ and additional background information $\boldsymbol{v}_i$. As noted above, initial observations are already shaped by unobserved individual characteristics, which Heckman's specification captures by including the individual specific effect ξ_i with a scale factor λ that allows for a different effect magnitude of unobserved characteristics on initial preferences.[9] Finally, ϵ_{i0} is a random disturbance term at the initial condition assumed uncorrelated with other errors, i.e. $\text{Cov}(\epsilon_{i0}, \epsilon_{it}) = 0, \ \forall t > 0$. Monte Carlo evidence indicates

[7] Models which ignore this problem and specify initial conditions as exogenous can lead to severely biased estimates of the most central parameters of a dynamic panel model, namely individual random effects and preference persistence (e.g. Heckman 1981b; Fotouhi 2005; Arulampalam and Stewart 2009).

[8] As Anderson and Hsiao (1981: 598) note, this is a problem specific the short dynamic panels (such as household or election panels), since on cannot credible assume that $T \to \infty$.

[9] It facilitates a simple specification test of the appropriateness of assuming independence of initial conditions and unobserved individual effects: this assumption is rejected if $\lambda \neq 0$. This parametrization is sometimes called a factor-analytic formulation of random effects (e.g Skrondal and Rabe-Hesketh 2004). Alternatively, one could introduce a second set of random effects with fixed variance in (6), and estimate the covariance between them and those in equation (3). The present formulation is somewhat more intuitive and allows for a more straightforward test of exogeneity by testing the parameter λ instead of a covariance.

that this approximation works well in short panels (Heckman 1981a; Akay 2011).[10] A somewhat more detailed discussion can be found in [online] appendix A.

Jointly estimating (1) – (4) and (6) yields a model that deals with four of the five central problems outlined in the introduction. The dynamic model is supposed to capture serial correlation of responses given at different points in time by the same individual (e.g. Beck and Katz 1996). An estimate of this correlation is given by ρ defined in equation (5). To test for remaining autocorrelation, latent residuals (Albert and Chib 1995) can be used. I calculate remaining residual correlation as:

$$\hat{r} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} \mu_{it} \mu_{it-1}}{\sum_{i=1}^{N} \sum_{t=2}^{T} \mu_{it}^2} \tag{7}$$

where μ_{it} stands for the linear predictor used in (3). If the specification succeeds in modeling individuals' correlated responses over time, $\hat{r}$ should be close to zero.

2.3. *Prior specifications*

Model specification is completed by assigning (hyper-) priors to all parameters.[11] Priors for intercept and parameters of individual characteristics, in both dynamics and initial condition equations are diffuse with mean zero and large variance to yield regression-type estimates:

$$\beta, \delta \sim N(0, 100). \tag{8}$$

I use a normal distributed prior for ϕ, the parameter capturing persistence of preferences. I set a prior mean of 0.5 indicating an *a priori* expectation that persistence is not zero, but use a very large variance to yield a diffuse prior:

$$\phi \sim N(0.5, 100). \tag{9}$$

More informative priors might be preferable in some applications, e.g. by restricting ϕ using an uniform prior on $U(-1, 1,)$.

My hyperprior for the variance of individual random effects is uniform on the standard

[10] Alternative approximations, such as Wooldridge (2005), would specify the distribution of $\xi_i | y_{i0}, \boldsymbol{x}_{it}$, i.e. simply include the first panel observation among the regressors. This approximation is computationally easier to implement than Heckman's solution, which explains its predominance in applied research. However, if one specifies preferences as latent constructs, the variable one would need for conditioning on (z_{i0}) is not observable (Pudney 2006: 8). As another disadvantage, this approximation usually works less well in short panels (Akay 2011).

[11] Note that, as in every Bayesian analysis, sensitivity analyses for values of the hyperparameters should be carried out. For an overview of robustness check strategies see Gill (2008a: 199f.). Basic regression-type priors can be checked by using different variances, For more specific or complex priors, I describe sensitivity check strategies in the text.

deviation, bounded between zero and 10. Gelman (2006) recommends this prior over the more commonly used inverse Gamma specification (Spiegelhalter et al. 1997).

$$\sqrt{\sigma_\xi^2} \sim U(0, 10). \tag{10}$$

However, this prior has the disadvantage of assigning equal probability to unrealistically large random effect variances. While this can be seen as representing very little *a-priori* knowledge, some researchers might prefer a more informed specification using inverse gamma priors

$$\sigma_\xi^{-2} \sim \Gamma(a_0, b_0) \tag{11}$$

with values for a_0 and b_0 chosen using knowledge or expectations of the variation of the individual specific effects. I provide examples of such an analysis in [online] appendix D.

An uninformative prior for the random effect scale-factor in the initial condition equation (6) is a normal distribution centered at zero and with large variance:

$$\lambda \sim N(0, 100). \tag{12}$$

To ensure that thresholds follow the monotonicity constraint given in (2), I specify thresholds recursively ensuring that each subsequent threshold is larger than the previous one by adding a positive value v_τ. This is achieved by drawing v_τ from a distribution with positive support such as an exponential distribution (cf. Jackman 2009).[12] The first threshold is normalized to zero for identification; in a model without overall intercept it can be drawn from a normal distribution centered at zero with large variance.

$$\tau_1 = 0 \tag{13}$$
$$\tau_c = \tau_{c-1} + v_\tau, \quad c = 2, \dots, C-1 \tag{14}$$
$$v_\tau \sim \mathrm{Exp}(1). \tag{15}$$

2.4. *Model extensions*

Given its hierarchical nature, the model can be extended straightforwardly to capture higher order nesting by adding random effects for the relevant grouping factor. For example, individuals nested within families (e.g. Winkelmann 2005) or regions j $(j = 1, \dots, J)$ can be modeled by extending (3) to

$$z_{ijt} = \phi z_{ijt-1} + \boldsymbol{\beta}' \boldsymbol{x}_{ijt} + \xi_i + \psi_j + \epsilon_{ijt}$$

[12] Here I use an exponential distribution with rate one, but other parametrization are possible depending on one's *a priori* expected distance between thresholds. My specification expects a distance of one, which is close to the difference observed in a simple ordered probit regression. An alternative strategy for an ordering constraint is to order thresholds at each step of the MCMC sampler.

where ξ_i is the individual specific effect, and ψ_j represents the regional random effect. Initial conditions are still modeled via (6). This is now a three level model with responses nested in individuals nested in regions. Region random effects are distributed $\psi_j \sim N(0, \sigma^2_\psi)$ with an appropriate hyperprior such as $\sigma_\psi \sim U(0, c)$.

3. ROBUST RANDOM EFFECTS

The discussion in the previous section assumed normally distributed random effects. This assumption goes almost unnoticed as it is standard in the vast majority of random effects or 'multilevel' models in the social sciences. However, assumptions about the distribution of individual random effects ξ_i are not innocuous and can have important *substantive* implications for panel data analysis.[13] When using a normal distribution as random effects prior, the well-known shrinkage property of hierarchical models (Gill 2008a: 183; Robert 2007: ch.10) pulls individuals with extreme ξ_i values towards one common mean. Multi-modality or interesting patterns of random effects might be obscured. Checks of the normality assumption can not be carried out using the already shrunken residuals (Kyung et al. 2010).

In this section I describe two strategies for a more robust estimation of individual heterogeneity: (1) accommodating more extreme individual random effects by specifying a distribution with heavier tails, such as a t-distribution with small degrees of freedom (Lange et al. 1989); (2) estimating the random effects distribution nonparametrically using Dirichlet process priors (e.g. Gill and Casella 2009).

3.1. *t-distributed random effects*

As an alternative to the normal distribution, a t distribution can be used as robust prior for random effects. A t distribution with small degrees of freedom has heavier tails and accommodates more extreme random effect values (cf. Lange et al. 1989; Gelman et al. 2004: ch.17). Thus, changing the distributional specification in (4) to

$$\xi_i \sim t(0, \sigma^2_\xi, \mathrm{df}) \tag{16}$$

yields a model with t-distributed random effects. However, estimating the degrees of freedom from the data – e.g. by assigning a uniform prior – is often rather difficult. For my goal of checking the robustness of the normal random effects assumption, choosing a small value, such as 4 degrees of freedom, is more appropriate (Gelman et al. 2004: 446).

[13] For a similar argument in the context of marketing models see Rossi et al. (2005: ch. 5); see Navarro et al. (2006) for experimental psychology.

3.2. *Dirichlet process random effects*

A more flexible alternative to assuming normally distributed random effects consists in estimating the random effects distribution non- or semi-parametrically. In the simpler linear dynamic panel case, a fixed effects approach can be employed without distributional assumptions – however this is unavailable for the current model (e.g. Nickell 1981; Heckman 1981b). Thus, when random effects have to be used, Arellano and Carrasco (2003) argue that (p. 126) "a semi-parametric random effects specification may represent a useful compromise" between the two.

In a frequentist framework, nonparametric estimation can be accomplished by using finite mixtures of normals or by approximating the random effects distribution by a finite number of mass points (e.g. Heckman and Singer 1984; Lindsay 1995; Aitkin 1999; Eckstein and Wolpin 1999; Vermunt 2004). When applied to substantive research questions, a central problem consists in how to choose the number of mixtures or mass points (Laird 1978; Follmann and Lambert 1989; Vermunt et al. 2008; Skrondal and Rabe-Hesketh 2004: 181f.).

In a fully Bayesian analysis, instead of assuming a distribution G for the random effects, one can place a Dirichlet process prior (Ferguson 1973, 1974) on G itself to indicate uncertainty about its shape (e.g. Kleinman and Ibrahim 1998; Gill and Casella 2009):

$$\xi_i \sim G \tag{17}$$

$$G \sim DP(\alpha, G_0) \tag{18}$$

A Dirichlet process is characterized by two components. The base distribution G_0 is the expectation of G – the distribution one would have used in a non-DP model (Escobar 1995: 98). In my current application this is the zero-centered normal distribution with estimated variance. The precision or dispersion parameter α determines the dispersion of the prior for G over its mean G_0 (Müller and Quintana 2004). Thus, using a Dirichlet process prior, each set of individual random effects $\{\xi_1, \ldots, \xi_N\}$ drawn from G lies in a set of K distinct values or 'subclusters' (with $K \leq N$) sampled from G_0: $\{\zeta_1, \ldots, \zeta_K\}$.[14] For each number of realized subclusters at any particular step of an MCMC sampler, random effects ξ_i are drawn from the set $\{\zeta_1, \ldots, \zeta_K\}$ via multinomial sampling. Define subcluster membership indicators $S = \{s_1, \ldots, s_N\}$ which are $s_i = k$ if $\xi_i = \zeta_k$; and $m_k = \#\{s_i = k\}$ as the number of random effects which share the same value ζ_k (i.e. they belong to the same subcluster k).[15]

To illustrate the working of the Dirichlet process, I describe the assignment of random effect ξ_i of a particular individual to a subcluster k, conditional on all remaining ran-

[14] The term "subcluster" is used to indicate that clustering is done nonparametrically and not based on substantive criteria (cf. Kyung et al. 2010)

[15] Thus, using a Dirichlet process prior provides discrete realizations from the infinite space of prior distributions with probability one (Ghosh and Ramamoorthi 2003; Müller and Quintana 2004). A more detailed discussion can be found in [online] appendix B.

dom effects $\xi_{[i]} = \{\xi_1, \ldots, \xi_{i-1}, \xi_{i+1}, \ldots, \xi_N\}$ being already assigned. Denote by $S_{[i]}$ the specific configuration of $N-1$ random effects into $K_{[i]}$ subclusters existing at this point, with $m_{[i],k} = \#\{s_i = k,\ k \neq i\}$ giving the number of individuals sharing a common value $\zeta_{[i],k}$. The conditional prior for ξ_i is (see Hanson et al. 2005 or Dunson et al. 2007 for details):

$$[\xi_i | \xi_{[i]}, K_{[i]}, S_{[i]}, \alpha] \sim \frac{\alpha}{\alpha+N-1} G_0 + \frac{1}{\alpha+N-1} \sum_{k \neq i} \delta(\xi_k) \tag{19}$$

$$\sim \frac{\alpha}{\alpha+N-1} G_0 + \frac{1}{\alpha+N-1} \sum_{k=1}^{K_{[i]}} m_{[i],k} \delta(\zeta_{[i],k}) \tag{20}$$

where $\delta(\cdot)$ now represents the Dirac delta function yielding a single value at its argument. In other words, ξ_i forms a new subcluster with probability $\alpha/\alpha + N - 1$, in which case it is drawn from G_0. Else, it gets value $\zeta_{[i],k}$ of an existing subcluster with multinomial probability according to $N_{[i],k}/\alpha + N - 1$. If one imagines a stream of individual random effects to be assigned, this leads to a *preferential attachment* clustering structure: as the number of individuals grows, the probability that a new individual is assigned to an already existing subcluster is proportional to the subcluster's size. The probability that a new individual forms a new subcluster of the Dirichlet Process is proportional to α, and if that happens, values for ξ_i are generated according to the base distribution G_0 (Müller et al. 2007).

The realized numbers of subclusters K is stochastic and is governed by α, which can be itself estimated from the data (see below). The role of α can be visualized by inspecting its relationship with the expected number of subclusters (Hanson et al. 2005), which can be approximated as (Antoniak 1974; Escobar 1995):

$$E(k|\alpha, n) \approx \alpha \log[(\alpha + N)/\alpha]. \tag{21}$$

Figure 1 plots the expected number of subclusters as a function of the number of individuals for different values of α. This nicely illustrates the logarithmic nature of the preferential attachment property of the Dirichlet process and conforms to intuitions about the relationship between the number of different subclusters and the number of individuals: As more and more individuals are observed, the chance of observing new and unexpected random effect values increases, but at a decreasing rate.

In the dynamic panel model with random effects, considered here, the set of parameters in the base distribution is simply $G_0 = \{p(\sigma_\xi^2)\}$ with a uniform hyperprior $\sigma_\xi \sim U(0, 10)$ as before. Thus the marginal distribution – averaging over all possible G – yields a mixture of normal distributions with the number of subclusters K randomly varying between 1 and N (see Kleinman and Ibrahim 1998 for a similar setup).[16] The individual specific random effect variance parameters are either selected from the $K_{[i]}$ existing values $\zeta_k = \sigma^2_{\xi,k}$ drawn from G_0,

[16] In practical implementations using a Truncated Dirichlet process, the number of subclusters is restricted to some truncation value $T \ll N$. See appendix B for details.

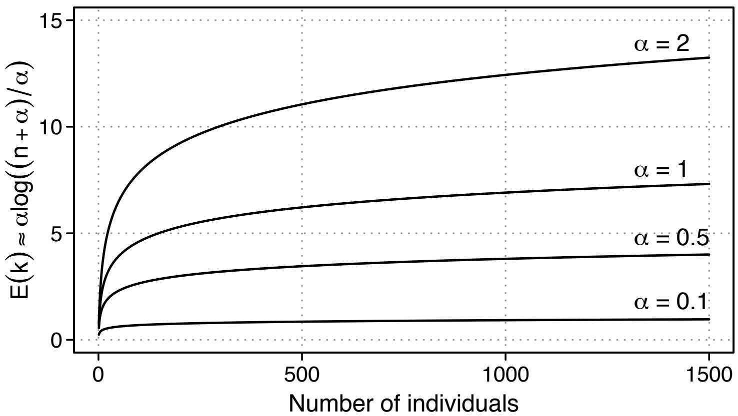

Figure 1: Expected number of subclusters as function of sample size and Dirichlet process precision parameter α

or created via a fresh draw from G_0. A more detailed technical discussion of the Dirichlet process and its implementation is available in [online] appendix B.

Estimating dispersion parameter α from the data

The dispersion parameter, α, is a central parameter of the model. Higher values of α increase not only the number of expected subclusters, but also the rate with which new ones are created by the Dirichlet Process. Given the absence of clear prior expectations about values of α, its value can be determined by the data yielding a mixture of Dirichlet processes (Antoniak 1974). In a fully Bayesian context this is achieved by assigning it a hyperprior:

$$\alpha \sim \Gamma(a_0, b_0). \tag{22}$$

The gamma distribution is a common choice for this problem (Escobar and West 1998; Jara et al. 2007), however its parameters do not allow for an intuitive prediction of its effect on the model.[17] Kottas et al. (2005) provide an approximation to the relationship between Γ-prior parameters and expectation and variance of the number of subclusters K, which can be used to choose semi-informed prior values (for more details see appendix C). I select parameters for the gamma hyperprior so that they yield 8 a priori expected clusters with a standard deviation of 4, which yields parameters $a_0 = 5.16$ and $b_0 = 4.54$ for the gamma prior. To

[17] Specifying an essentially flat prior for computational reasons is common in political science applications (Jackman 2000; but see Jackman and Western 1994), but is of somewhat questionable value here. Even medium-sized values of α lead to a large number of clusters, which in the limiting case creates one cluster per individual – essentially defying the purpose of the hierarchical setup. Therefore, I argue to use a semi-informed prior specification (Gill and Casella 2009: 3) for the DP precision parameter. Kyung et al. (2010) provide alternative strategies of sampling the concentration parameter.

check the sensitivity of this specification, I also used values which lead to a prior expectation of half the number of clusters ($a_0 = 0.921$ and $b_0 = 1.435$). In an alternative strategy (and robustness test), one can forgo estimation of α and instead fix it to a set of pre-specified values, e.g. $\alpha = \{0.5, 1, 2, 10\}$, in order to determine the robustness of one's estimates to increasingly larger numbers of random effects subclusters. The approximations given in equation (21) and Figure 1 can serve as guidelines relating values of α to expected subclustering and one's sample size.

4. APPLICATION: DYNAMIC PREFERENCES FOR REDISTRIBUTION

A recent wave of research in (comparative) political economy has augmented macro-level studies of redistribution by concentrating on individual-level factors influencing redistribution preferences (see, among many, Moene and Wallerstein 2001; Iversen and Soskice 2001; Alesina and La Ferrara 2005; Alesina and Angeletos 2005; Cusack et al. 2005; Scheve and Stasavage 2006; Shayo 2009; Rehm 2011; Rehm et al. 2012). Studies examining preferences for redistribution and government intervention in the economy are usually cross-sectional and ignore dynamic aspects of preference formation.[18] As a consequence, estimates of key variables, such as the effect of job loss (as in Cusack et al. 2008) might be influenced by unobserved factors, such as ability and motivation, as well as by persistent preferences.[19]

In this section, I present a short study of the dynamics of individual redistribution preferences, by applying the model outlined before to repeated measurements of individuals' preferred level of government intervention. More specifically, I examine individual responses to the question if government has the obligation to provide jobs. This survey item correlates highly with other widely used measure of general redistribution preferences.[20] I examine the effects of income and wealth and of 'socio-economic shocks' such as becoming unemployed or getting divorced. For a recent summary of the theoretical relevance of these factors see Alesina and Giuliano (2011).

4.1. *Data and variables*

I use data from the British Household Panel Survey, conducted between 1991 and 2008, which provides measurements of my dependent variable on 7 occasions. I use the original ('Essex') sample and create a balanced panel using individuals who provide responses to all seven waves.[21] This provides me with data on 1958 individuals observed over a span of 17 years.

[18] But see recent research based on experimental evidence, e.g. Margalit (2011), Neustadt (2010).

[19] This should not be read as a critique of this particular paper, given that the authors' interest lies in a comparative analysis (where panel data is unavailable).

[20] Its correlation with a latent preference measure of several redistribution items (following the methodology of Stegmueller 2011) using data for the UK from the International Social Survey Programme is 0.64.

[21] Items are available in waves A, C, E, G, J, N, and Q. Estimating the model using multiple imputation for missing values provides results that are substantively similar to the ones presented here, as does an analysis

Responses to the item "It is the government's responsibility to provide a job for everyone who wants one" are captured using the usual 5 point strongly agree – strongly disagree scale.[22] Since both extreme ends of the response categories are rather sparsely populated, I combine categories to yield a clear three-category response vector, which indicates if preferred levels government activity should stay the same (0), or should be increased (1) or decreased (−1).[23] Thus, the relationship between observed responses and the latent preference variable is given by:

$$y_{it} = \begin{cases} -1 & \text{if } \; z_{it} < \tau_1 = 0 \\ 0 & \text{if } \; \tau_1 = 0 < z_{it} < \tau_2 \\ 1 & \text{if } \; \tau_2 < z_{it} \; . \end{cases}$$

Income is captured by both household income, and the share of a respondent's income of total household income. I measure income as real equivalent household income, i.e., it is deflated using the consumer price index with base year 2005 and adjusted for household size using the modified OECD equivalence scale (Hagenaars et al. 1994). I decompose income into a time varying and a time constant part. Thus, I estimate both a level and a shock effect, which mirrors the theoretical idea of permanent and transient income components (Friedman 1957). More precisely, observed income w_{it} is decomposed as $w_{it} = \bar{w}_i + (w_{it} - \bar{w}_i)$ with appropriately specified regression weights for both terms. Household wealth is captured by the estimated value of a respondent's house. Definitions and descriptive statistics of all other independent variables used in the analysis can be found in Table 1. Following Gelman (2008), in all models estimated below I centered and scaled all continuous variables by dividing by two standard deviations (which makes them roughly comparable to binary covariates).

4.2. *Results*

First, I describe results obtained from estimating the model described in section 2 assuming normally distributed random effects. I use a 66% subsample of individuals from the full sample. Results are obtained by MCMC sampling using two chains run for 500,000 iterations thinned by a factor of 25. 200,000 previous iterations are discarded as burn-in. The model is implemented using JAGS (version 3.1.0) with a truncation threshold of 20 (see the discussion of the Truncated Dirichlet Process in appendix B).[24] Diagnostics suggested by Brooks and Roberts (1998) and Gelman and Rubin (1992) do not show signs of absence of 'convergence'.[25]

which uses an unbalanced panel of respondents who participated in at least three waves.

[22] Categories are labeled strongly agree; agree; neither agree nor disagree; disagree; strongly disagree.

[23] Note that in single index models, such as this one, consistency of the estimates is not hampered by combining categories. See Franses and Cramer (2010) for a further discussion on combining categories in ordered response models. Furthermore, this dependent variable clearly represents a situation where linear models are not appropriate.

[24] A second run with a truncation value of 40 yields a maximum posterior sampled value for K of 17, which indicates that a truncation level of $T = 20$ was appropriate (see [online] appendix B).

[25] The posterior samples converge early, but I ran the sampler for longer, providing more draws for the thresholds in order to avoid non-convergence in this part of the model (cf. Gill 2008b). I conducted an "insurance run"

Table 1: Descriptive statistics of independent variables

Name	Description	mean	sd
Income	Equivalent household income (in 10,000 £)		
Permanent	Permanent income component	4.590	2.390
Transitory	Transitory income component	0.000	2.558
Income share	R's share of total HH income	0.532	0.300
House value	Estimated house value (in 100,000 £)	1.181	1.397
Owner	House owned outright or with mortgage	0.819	0.385
Unemployed	Unemployed	0.033	0.178
Union member	Union member	0.231	0.421
Divorced	Divorced	0.059	0.235
HH size	Size of Household	3.202	1.272
N kids	Number of kids in HH	0.906	1.084
Female	Gender: female	0.511	0.500
Age	Age in years	3.963	0.960
Nonwhite[†]	Ethnic group non-white	0.032	0.175
Education[b,†]			
Degree	University degree	0.199	0.399
A-levels	A level or higher national diploma	0.193	0.394
O-levels	O level or GCSE	0.411	0.492
London[†]	R grew up in greater London area	0.100	0.300
Parents' jobs[c,†]	Parents' job status		
unskilled	Blue collar, unskilled jobs	0.163	0.369
skilled	Blue collar, skilled jobs	0.223	0.416
white-collar	White collar	0.145	0.352
self-employed	Self-employed	0.144	0.351
N rows	13706		
N individuals	1958		

[†] Variables are time constant
[a] Equivalized using OECD scale; deflated using consumer price index, 2005 prices
[b] Reference category: no or primary education
[c] Reference category: Managers, Salariat; dominance coding

Resulting estimates are shown in Table 2, where I provide posterior means and standard deviations as well as 95% highest posterior density regions. Concentrating on central dynamic parameters, I find a significant amount of preference persistence: ϕ is estimated as 0.23 with a small posterior standard deviation. An estimated random effect variance, σ^2_ξ, of 0.83±0.09 underscores the importance of controlling for unobserved individual heterogeneity. The proportion of the total variance that is due to unobserved individual factors, ρ, is estimated as 45±3%. Thus, almost half of the difference in preferences between individuals is due to unobserved factors such as ability or motivation (which remains hidden in cross-sectional studies). Clearly, more research is needed to capture such unobserved individual characteristics.

As described in section 2, a specification test for the independence of initial conditions and unobserved individual effects is obtained by testing if λ is equal to zero. This is clearly rejected by an estimate of 1.19 and a HPD region far away from zero. In other words, initial conditions should be modeled as endogenous to individual (observed an unobserved) characteristics. Relevant covariates in the initial conditions equation are age, income, and notably education, as well as pre-sample information on parental background. For example, individuals who grew up in a working class household already have substantively higher preferences for government intervention at the start of the panel.

(Gill 2008b: 173): running the sampler for twice as many iterations. Estimates for all key model parameters are virtually identical; with the largest difference being 0.0019. All code and diagnostics are available in the author's dataverse.

Table 2: Posterior summary for Hierarchical dynamic latent ordered probit model.

Initial conditions		Mean	SD	95% HPD		Dynamics		Mean	SD	95% HPD	
δ_1	[Permanent income]	−0.422	0.132	−0.679	−0.162	β_1	[Permanent income]	−0.404	0.071	−0.546	−0.267
δ_2	[Transitory income]	−0.033	0.122	−0.273	0.202	β_2	[Transitory income]	−0.059	0.038	−0.133	0.018
δ_3	[R's Income share]	−0.015	0.107	−0.225	0.194	β_3	[R's Income share]	−0.113	0.051	−0.214	−0.016
δ_4	[House value]	−0.219	0.228	−0.656	0.234	β_4	[House value]	−0.115	0.052	−0.220	−0.016
δ_5	[House owner]	−0.073	0.102	−0.267	0.133	β_5	[House owner]	−0.026	0.047	−0.117	0.067
δ_6	[HH size]	0.229	0.185	−0.143	0.579	β_6	[HH size]	0.085	0.079	−0.072	0.237
δ_7	[N kids in HH]	−0.047	0.168	−0.381	0.278	β_7	[N kids in HH]	−0.021	0.065	−0.151	0.105
δ_8	[Union member]	0.155	0.100	−0.045	0.349	β_8	[Union member]	0.148	0.049	0.054	0.244
δ_9	[Age]	−0.551	0.115	−0.779	−0.326	β_9	[Age]	−0.314	0.056	−0.424	−0.202
δ_{10}	[Female]	0.271	0.090	0.094	0.446	β_{10}	[Female]	0.235	0.064	0.112	0.360
δ_{11}	[Divorced]	0.124	0.237	−0.340	0.589	β_{11}	[Divorced]	0.058	0.087	−0.111	0.225
δ_{12}	[Unemployed]	−0.398	0.216	−0.815	0.030	β_{12}	[Unemployed]	0.140	0.100	−0.055	0.334
δ_{13}	[Non-white]	0.317	0.282	−0.231	0.871	β_{13}	[Non-white]	0.606	0.171	0.272	0.941
δ_{14}	[Degree]	−0.524	0.175	−0.866	−0.181	β_{14}	[Degree]	−0.669	0.105	−0.879	−0.469
δ_{15}	[A-levels]	−0.521	0.169	−0.849	−0.185	β_{15}	[A-levels]	−0.465	0.102	−0.671	−0.269
δ_{16}	[O-levels]	−0.431	0.143	−0.712	−0.155	β_{16}	[O-levels]	−0.321	0.085	−0.493	−0.158
δ_{17}	[Parents: unskilled]	0.464	0.138	0.189	0.732	β_0	[Intercept]	0.213	0.043	0.131	0.299
δ_{18}	[Parents: skilled]	0.233	0.117	0.004	0.462	τ_2		0.818	0.019	0.781	0.855
δ_{19}	[Parents: white-collar]	0.299	0.125	0.055	0.544	ϕ		0.226	0.026	0.176	0.277
δ_{20}	[Parents: self-empl.]	0.115	0.130	−0.137	0.373	σ^2_ξ		0.829	0.092	0.654	1.013
δ_{21}	[London]	0.047	0.139	−0.220	0.326	$\hat{r}$		−0.009	0.012	−0.031	0.015
λ		1.185	0.098	0.995	1.379	ρ		0.452	0.027	0.398	0.506
Deviance						25705					
DIC						26259					
Posterior predictive p-value						0.528					

Note: Based on 20.000 MCMC draws. Threshold τ_1 fixed at 0. Balanced panel, 9044 rows, 1292 individuals. Posterior predictive check calculated for mean number of preference changes.

In a dynamic panel model a central quantity of interest are long-run or steady-state relationships between z and $\boldsymbol{x}$ taking preference persistence into account. Since I fixed the scale of the error variance to 1, steady-state effects are calculated as $\beta/(1-\phi)$. Using 5000 draws from the relevant parameters' posterior distributions, I calculate posterior means and standard deviations of steady-state effects, displayed in Table 3. For easier interpretation, I provide them both in the metric of the latent dependent variable z, and calculated as first differences in predicted probabilities of preferring more government intervention resulting from a unit-change in a covariate. For discrete variables this reflects a change from 0 to 1; for continuous variables this represents a change of 2 standard deviations (cf. Gelman 2008).

Long-run estimates of wealth captured by permanent income and house value show a strong and negative relationship with preferences for government intervention. All else equal, a unit-change of permanent income reduces an individual's probability to opt for more government intervention by 17±3 percentage points. It is noteworthy that income shocks have little effect on preferences and are statistically indistinguishable from zero. I find the same for the estimated long-run effect of becoming unemployed, which is large but has a posterior density that includes zero. This also holds for its parameter estimates displayed in Table 2. Excluding all income effects from the model does not change this finding. This points to the relevance of including preference persistence and (especially) unobserved individual heterogeneity in studies of individual preferences. It is this specification of unobserved heterogeneity which I turn to next.

4.3. *Robust random effects results*

To check the robustness of my random effects specification, I re-estimated my model using the strategies outlined in section 3. A model with t-distributed random effects with 4 degrees of freedom produces a lower estimate of the random effects variance, σ_{ξ}^2, of 0.77 with a 95% HPD region ranging from 0.68 to 0.86. However, all other model parameters, including the preference persistence parameter ϕ, are estimated at virtually the same values (at 2 sf.). When using a more flexible density estimate of the random effects distribution using a Dirichlet process prior, more differences emerge.[26]

Figure 2 plots a kernel density estimate of the distribution of random effect estimates (more precisely their posterior expectation) from the Dirichlet process hierarchical model. Clearly the distribution of random effects differs from the traditionally made normal assumption, being slightly skewed and more peaked. However, there is no clear evidence of multi-modality or the existence of extreme random effects in the tails of the distribution. This suggest that central model parameters might not be too strongly affected by differences in random effect estimates.

To illustrate differences in parameter estimates that emerge when using different random

[26] A full table of parameter estimates for the DP prior model is given in appendix E.

Table 3: Steady-state effects. Calculated on the scale of the latent variable z and as predicted probability of responding in the highest category. Posterior means and standard deviations.

	z-metric		$P(y_i = 1)$	
	Mean	SD	Mean	SD
Permanent income	−0.521	0.093	−0.172	0.031
Transitory income	−0.076	0.050	−0.025	0.016
R's Income share	−0.146	0.066	−0.049	0.022
House value	−0.149	0.068	−0.049	0.022
House owner	−0.034	0.060	−0.014	0.026
Household size	0.110	0.103	0.036	0.034
N kids in HH	−0.027	0.085	−0.009	0.028
Union member	0.192	0.064	0.065	0.022
Age	−0.406	0.074	−0.134	0.024
Female	0.304	0.083	0.100	0.027
Divorced	0.074	0.111	0.027	0.039
Unemployed	0.178	0.131	0.064	0.046
Non-white	0.785	0.221	0.294	0.085
Degree	−0.866	0.138	−0.233	0.030
A-levels	−0.601	0.134	−0.174	0.034
O-levels	−0.416	0.111	−0.134	0.035

Note: Calculated using 5000 simulated values. Predicted probabilities represent unit-change in variable holding all else constant.

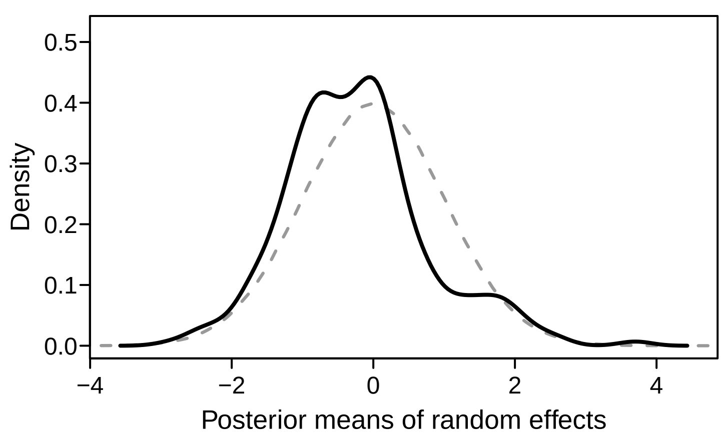

Figure 2: Distribution of random effects using a Dirichlet process prior. Density estimate of posterior means of random effects ξ_i evaluated over grid of 200 points. Normal distribution (- - -) shown for comparison.

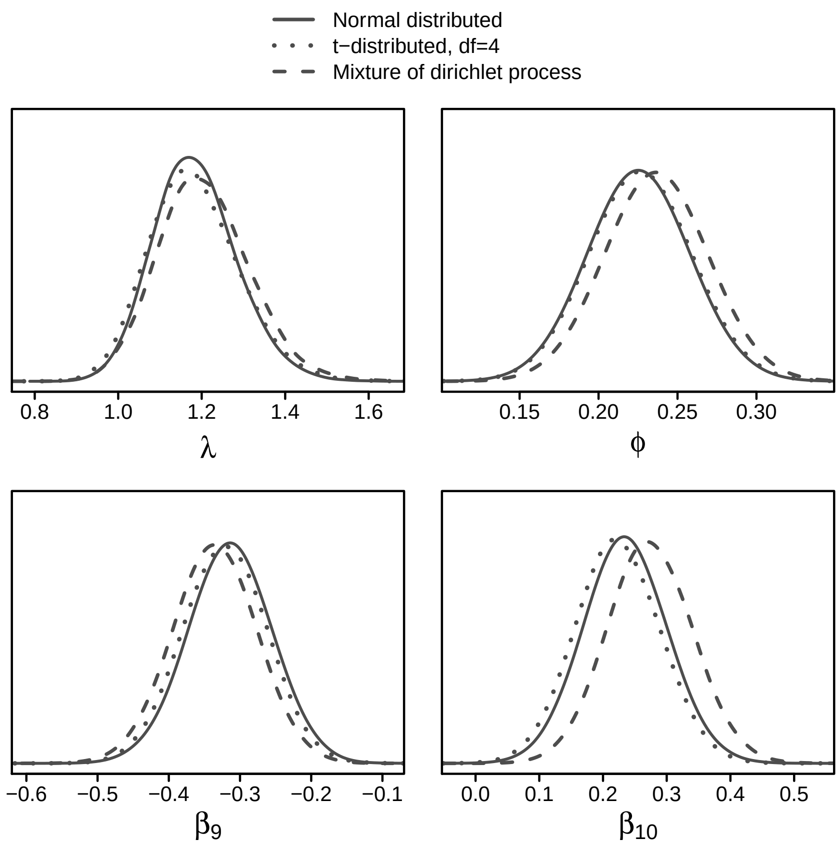

Figure 3: Consequences of different random effect prior specifications. Posterior distributions of selected parameters obtained using a normal distribution; a *t* distribution with 4 df.; and mixture of Dirichlet processes as prior.

Table 4: Steady state estimates from Dirichlet random effects model. Panel (A) shows estimated steady state effects. Panel (B) shows difference to normal random effects model. Posterior means and standard deviations.

	(A) Estimates				(B) Difference to normal RE			
	z-metric		$P(y_i = 1)$		z-metric		$P(y_i = 1)$	
Permanent inc.	−0.516	0.088	−0.187	0.031	0.005	0.130†	−0.015	0.043†
Transitory inc.	−0.081	0.051	−0.030	0.019	−0.006	0.071†	−0.004	0.025†
Income share	−0.137	0.065	−0.050	0.024	0.007	0.091†	−0.002	0.033†
House value	−0.145	0.068	−0.053	0.025	0.004	0.097†	−0.003	0.033†
House owner	−0.025	0.062	−0.013	0.029	0.007	0.086†	0.002	0.039†
Household size	0.142	0.105	0.052	0.038	0.034	0.147†	0.016	0.051†
N kids in HH	−0.041	0.087	−0.015	0.031	−0.014	0.120†	−0.006	0.042†
Union member	0.193	0.063	0.072	0.024	0.002	0.089†	0.007	0.033†
Age	−0.440	0.075	−0.160	0.028	−0.033	0.105†	−0.026	0.037†
Female	0.359	0.086	0.131	0.031	0.054	0.119†	0.031	0.039†
Divorced	0.077	0.112	0.028	0.042	0.002	0.157†	0.002	0.057†
Unemployed	0.181	0.129	0.069	0.050	0.000	0.180†	0.004	0.068†
Non-white	0.635	0.224	0.245	0.084	−0.145	0.315†	−0.051	0.119†
Degree	−0.844	0.134	−0.264	0.037	0.017	0.196†	−0.031	0.045†
A-levels	−0.614	0.134	−0.202	0.040	−0.016	0.189†	−0.029	0.051†
O-levels	−0.422	0.109	−0.152	0.038	−0.006	0.152†	−0.018	0.051†

Note: Calculated using 5000 simulated values. Predicted probabilities represent unit-change in variable holding all else constant. Differences in panel (B) calculated as DP random effects estimates − normal random effects estimates. Difference estimates whose 95% HPD interval includes zero are marked by †.

effect prior specifications, I plot posterior distributions of some selected parameters in Figure 3. It shows posterior distributions of the initial condition random effects scale factor λ, preference persistence ϕ, and estimates of age and being female obtained using normal, t, and DP random effects. Estimates of the scale factor and preference persistence are indistinguishable between normal and t-distributed random effects. However, they are larger under the DP prior specification, especially for preference persistence. Results for substantive covariates also differ when using a flexible DP prior specification. My estimate for the influence of age on preferences for government intervention becomes smaller, while the posterior distribution of being female is clearly shifted to the right, indicating an even stronger effect. Nonetheless, the magnitude of these differences is limited and other covariate estimates are somewhat less affected than the ones shown here.

To assess if these differences change one's substantive results, it is advisable to focus again on steady state estimates calculated from the model. In Table 4, panel (A), I provide steady state effects from the DP random effects model. As before, I calculate them in the metric of the latent variable and as predicted probabilities of preferring more government involvement. In panel (B) I calculate the difference to the steady state estimates based on normal random effects, shown in Table 3, and mark difference estimates whose 95% HPD interval contains zero by †. I find that differences are especially marked for estimates of time-constants covariates. The difference between the estimated effect of holding an advanced degree is almost three percentage points, while the effect of being non-white differs by 5 percentage points. However, when taking uncertainties of my estimates into account, these differences appear to not be statistically relevant: in each and every case the 95% highest posterior density interval of the difference contains zero. Thus, in this particular application, one can conclude that substantive results obtained with a 'simple' gaussian random effects specification are robust to violations of the distributional assumption of unobserved individual heterogeneity.

5. CONCLUSION

Central aim of this paper is to present a modeling strategy for analyzing the dynamics of individual preferences or attitudes using panel data. I employ the idea of an underlying latent continuous variable, which generates observed categorical preference measures. The dynamics of the model are also specified on the level of the latent variable, since it should be one's latent past preference – not observed survey scores – providing feedback to current preferences. Furthermore, I explicitly model initial conditions, following the approach suggested by Heckman (1981a, b). I capture unobserved individual heterogeneity using random effects and discuss possible shortcomings of the usual distributional assumptions. I employ a distinctively 'Bayesian' solution to this problem, which is to specify a prior over possible random effect distributions, in order to capture uncertainty about its true form. This yields flexible nonparametric density estimation of random effects, which I use to assess the robustness of my findings.

Applying the model to data on individuals' preferences for government intervention over a span of 17 years, clearly shows the necessity of employing a Hierarchical dynamic panel modeling approach. First, I find a significant level of preference persistence. In other words, individuals' preferences are 'sticky', and covariate estimates will be biased when ignoring this fact. Second, initial conditions matter. Individuals enter the panel study with preferences already shaped by pre-sample variables and observed and unobserved characteristics. Third, nearly half of the total variation in preferences is due to unobserved individual factors, such as motivation or ability. Using both parametric and semi-parametric random effects specifications, I show that these findings are robust to distributional assumptions.

Existing political science research on individual preferences and attitudes using cross-sectional data should be augmented into the time domain to explicitly study dynamic implications of theories. Using panel data and an appropriate dynamic model provides the tools to generate new insights into how individual preferences evolve over time, how they are shaped by observed and unobserved individual characteristics, and how individuals adjust their preferences in reaction to socio-economic shocks.

REFERENCES

Aitkin, Murray. 1999. A general maximum likelihood analysis of variance components in generalized linear models. *Biometrics* 55:117–128.

Akay, Alpaslan. 2011. Finite-sample comparison of alternative methods for estimating dynamic panel data models. *Journal of Applied Econometrics (preprint)* .

Albert, James H and Siddhartha Chib. 1993. Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* 88:669–679.

Albert, Jim and Siddhartha Chib. 1995. Bayesian Residual Analysis for Binary Response Regression Models. *Biometrika* 82:747–759.

Alesina, Alberto and George-Marios Angeletos. 2005. Fairness and Redistribution. *American Economic Review* 95:960–980.

Alesina, Alberto and Paola Giuliano. 2011. Preferences for Redistribution. In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, and Matthew O. Jackson. San Diego: North-Holland, 93–131.

Alesina, Alberto and Eliana La Ferrara. 2005. Preferences for redistribution in the land of opportunities. *Journal of Public Economics* 89:897–931.

Anderson, T W and C Hsiao. 1981. Estimation of dynamic models with error components. *Journal of the American Statistical Association* 76:598–606.

Antoniak, Charles E. 1974. Mixtures of Dirichlet Processes with Applications to Bayesian Nonparametric Problems. *The Annals of Statistics* 2:1152–1174.

Arellano, Manuel and Stephen Bond. 1991. Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations. *The Review of Economic Studies* 58:277.

Arellano, Manuel and Raquel Carrasco. 2003. Binary choice panel data models with predetermined variables. *Journal of Econometrics* 115:125–157.

Arulampalam, W. 2000. Unemployment persistence. *Oxford Economic Papers* 52:24–50.

Arulampalam, Wiji and Mark B. Stewart. 2009. Simplified Implementation of the Heckman Estimator of the Dynamic Probit Model and a Comparison with Alternative Estimators. *Oxford Bulletin of Economics and Statistics* 71:659–681.

Bartels, Brandon L, Janet M Box-steffensmeier, Corwin D Smidt, and Rene M Smith. 2011. The dynamic properties of individual-level party identification in the United States. *Electoral Studies* 30:210–222.

Bartels, Larry M. 1999. Panel effects in the American national election studies. *Political Analysis* 8:1–20.

Beck, Nathaniel and Jonathan N Katz. 1996. Nuisance vs. Substance: Specifying and Estimating Time-Series-Cross-Section Models. *Political Analysis* 6:1–36.

Blackwell, David and James B MacQueen. 1973. Ferguson Distributions Via Polya Urn Schemes. *The Annals of Statistics* 1:353–355.

Blundell, Richard and Stephen Bond. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87:115–143.

Brooks, Stephen P and Gareth O Roberts. 1998. Convergence assessment techniques for Markov chain Monte Carlo. *Statistics and Computing* 8:319–335.

Cusack, Thomas, Torben Iversen, and Phillip Rehm. 2008. Economic Shocks, Inequality, and Popular Support for Redistribution. In *Democracy, Inequality, and Representation: A Comparative Perspective*, edited by Pablo Beramendi and Christopher J Anderson. New York: Russell Sage Foundation, 203–231.

Cusack, Thomas, Torbern Iversen, and Phillip Rehm. 2005. Risks At Work: The Demand And Supply Sides Of Government Redistribution. *Oxford Review Of Economic Policy* 22:365–389.

Czado, Claudia, Anette Heyn, and Gernot Müller. 2011. Modeling individual migraine severity with autoregressive ordered probit models. *Statistical Methods and Application* 20:101–121.

Dunson, David B, Natesh Pillai, and Ju-Hyun Park. 2007. Bayesian density regression. *Journal of the Royal Statistical Society B* 69:163–183.

Eckstein, Zvi and Kenneth Wolpin. 1999. Why Youths Drop Out Of High School: The Impact Of Preferences, Opportunities, And Abilities. *Econometrica* 67:1295–1339.

Escobar, Michael D. 1995. Nonparametric Bayesian methods in hierarchical models. *Journal of Statistical Planning and Inference* 43:97–106.

Escobar, Michael D and Mike West. 1995. Bayesian Density Estimation and Inference Using Mixtures. *Journal of the American Statistical Association* 90:577–588.

Escobar, Michael D and Mike West. 1998. Computing Bayesian Nonparametric Hierarchical Models. In *Practical Nonparametric and Semiparametric Bayesian Statistics*, edited by Dipak K Dey, Peter Müller, and Debajyoti Sinha. Springer, 1–22.

Ferguson, Thomas S. 1973. A Bayesian Analysis of Some Nonparametric Problems. *The Annals of Statistics* 1:209–230.

Ferguson, Thomas S. 1974. Prior Distributions on Spaces of Probability Measures. *The Annals of Statistics* 2:615–629.

Follmann, Dean A and Diane Lambert. 1989. Generalizing Logistic Regression by Nonparametric Mixing. *Journal of the American Statistical Association* 84:295–300.

Fotouhi, Ali Reza. 2005. The initial conditions problem in longitudinal binary process: A simulation study. *Simulation Modelling Practice and Theory* 13:566–583.

Franses, Philip Hans and J.S. Cramer. 2010. On the number of categories in an ordered regression model. *Statistica Neerlandica* 64:125–128.

Friedman, M. 1957. *A Theory of the Consumption Function*. Princeton: Princeton University Press.

Gelman, Andrew. 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1:515–534.

Gelman, Andrew. 2008. Scaling regression inputs by dividing by two standard deviations. *Statistics in Medicine* :2865–2873.

Gelman, Andrew, John B Carlin, Hal S Stern, and Donald B Rubin. 2004. *Bayesian Data Analysis*. Boca Raton: Chapman & Hall.

Gelman, Andrew and Donald Rubin. 1992. Inference from Iterative Simulation Using Multiple Sequences. *Statistical Science* 7:457–511.

Ghosh, J K and R V Ramamoorthi. 2003. *Bayesian Nonparametrics*. New York: Springer.

Gill, Jeff. 2008a. *Bayesian Methods. A Social and Behavioral Sciences Approach*. Boca Raton: Chapman & Hall.

Gill, Jeff. 2008b. Is Partial-Dimension Convergence a Problem for Inferences from MCMC Algorithms? *Political Analysis* 16:153–178.

Gill, Jeff and George Casella. 2009. Nonparametric Priors for Ordinal Bayesian Social Science Models: Specification and Estimation. *Journal of the American Statistical Association* 104:1–12.

Greene, William and David Hensher. 2010. *Modeling Ordered Choices: A Primer*. Cambridge: Cambridge University Press.

Grimmer, Justin. 2010. An Introduction to Bayesian Inference via Variational Approximations. *Political Analysis* 19:32–47.

Hagenaars, A., K. de Vos, and M.A. Zaidi. 1994. Poverty Statistics in the Late 1980s: Research Based on Micro-data. Report, Office for Official Publications of the European Communities.

Hanson, Timothy E., Adam J Branscum, and Wesley O Johnson. 2005. Bayesian Nonparametric Modeling and Data Analysis: An Introduction. In *Handbook of Statistics, Vol. 25*. Elsevier, 245–278.

Harris, Mark N, Laszlo Matyas, and Patrick Sevestre. 2008. Dynamic Models for Short Panels. In *The Econometrics of Panel Data. Fundamentals and Recent Developments in Theory and Practice*, edited by Laszlo Matyas and Patrick Sevestre. Berlin: Springer, 249–278.

Hasegawa, Hikaru. 2009. Bayesian Dynamic Panel-Ordered Probit Model and Its Application to Subjective Well-Being. *Communications in Statistics - Simulation and Computation* 38:1321–1347.

Heckman, James J. 1978. Dummy Endogeneous Variables in a Simultaneous Equation System. *Econometrica* 46:931–959.

Heckman, James J. 1981a. Heterogeneity and State Dependence. In *Studies in Labor Markets*, edited by Sherwin Rosen. Chicago: University of Chicago Press, 91–140.

Heckman, James J. 1981b. The incidental parameters problem and the problem of initial conditions in estimating a discrete time-discrete data stochastic process. In *Structural Analysis of Discrete Data with Econometric Applications*, edited by C. F. Manski and Daniel McFadden. Cambridge: MIT Press, 179–195.

Heckman, James J and B. Singer. 1984. A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data. *Econometrica* 52:271–320.

Imai, Kosuke, Ying Lu, and Aaron Strauss. 2008. Bayesian and Likelihood Inference for 2 x 2 Ecological Tables: An Incomplete-Data Approach. *Political Analysis* 16:41–69.

Ishwaran, Hemant and Lancelot F James. 2001. Gibbs Sampling Methods for Stick-Breaking Priors. *Journal of the American Statistical Association* 96:161–173.

Ishwaran, Hemant and Mahmoud Zarepour. 2000. Markov chain Monte Carlo in approximate Dirichlet and beta two-parameter process hierarchical models. *Biometrika* 87:371–390.

Ishwaran, Hemant and Mahmoud Zarepour. 2002. Exact and approximate sum representations for the Dirichlet process. *Canadian Journal Of Statistics* 30:269–283.

Iversen, Torben and David Soskice. 2001. An Asset Theory of Social Policy Preferences. *American Political Science Review* 95:875–893.

Jackman, Simon. 2000. Estimation and Inference are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation. *Political Analysis* 8:307–332.

Jackman, Simon and Bruce Western. 1994. Bayesian Inference for Comparative Research. *American Political Science Review* 88:412–423.

Jackman, Simon D. 2009. *Bayesian Analysis for the Social Sciences*. New York: Wiley.

Jara, Alejandro, Maria Jose Garcia-Zattera, and Emmanuel Lesaffre. 2007. A Dirichlet process mixture model for the analysis of correlated binary responses. *Computational Statistics and Data Analysis* 51:5402–5415.

Johnson, Valen E and Jim H Albert. 1999. *Ordinal Data Modeling*. New York: Springer.

Keane, MP. 1997. Modeling heterogeneity and state dependence in consumer choice behavior. *Journal of Business & Economic Statistics* 15:310–327.

Kleinman, Ken P. and Joseph G. Ibrahim. 1998. A Semiparametric Bayesian Approach to the Random Effects Model. *Biometrics* 54:921–938.

Kottas, Athanasios, Peter Müller, and Fernando Quintana. 2005. Nonparametric Bayesian Modeling for Multivariate Ordinal Data. *Journal of Computational and Graphical Statistics* 14:610–625.

Kyung, Minjung, Jeff Gill, and George Casella. 2011. Sampling schemes for generalized linear Dirichlet process random effects models. *Statistical Methods & Applications* 20:259–290.

Kyung, Minyung, Jeff Gill, and George Casella. 2010. Estimation in Dirichlet Random Effects Models. *The Annals of Statistics* 38:979–1009.

Laird, N. 1978. Nonparametric maximum likelihood estimation of a mixture distribution. *Journal of the American Statistical Association* 73:805–811.

Lange, Kenneth L, Roderick J A Little, and Jeremy M G Taylor. 1989. Robust Statistical Modeling Using the t-Distribution. *Journal of the American Statistical Association* 84:881–896.

Lindsay, Bruce. 1995. *Mixture Models: Theory, Geometry and Applications*. Hayward: Institute of Mathematical Statistics.

Liu, Jun S. 1996. Nonparametric Hierarchical Bayes Via Sequential Imputations. *The Annals of Statistics* 24:911–930.

Lo, Albert Y. 1984. On a class of Bayesian nonparametric estimates: I. Density estimates. *The Annals of Statistics* 12:351–357.

Lunn, David J., Jon Wakefield, and Amy Racine-Poon. 2001. Cumulative logit models for ordinal data: a case study involving allergic rhinitis severity scores. *Statistics in Medicine* 20:2261–2285.

MacEachern, S.N. and Peter Müller. 1998. Estimating mixture of Dirichlet process models. *Journal of Computational and Graphical Statistics* :223–238.

Margalit, Yotam. 2011. Ideology all the way down? The Impact of the Financial Crisis on Individuals' Welfare Policy Preferences. Paper presented at political economy workshop, Merton College, University of Oxford, April 2011.

McKelvey, Richard D and William Zavoina. 1975. A Statistical Model for the Analysis of Ordinal Level Dependent Variables. *Journal of Mathematical Sociology* 4:103–120.

Moene, Kark Ove and Michael Wallerstein. 2001. Inequality, Social Insurance and Redistribution. *American Political Science Review* 95:859–874.

Müller, Gernot and Claudia Czado. 2005. An Autoregressive Ordered Probit Model with Application to High-Frequency Financial Data. *Journal of Computational and Graphical Statistics* 14:320–338.

Müller, Peter and Fernando Quintana. 2004. Nonparametric Bayesian Data Analysis. *Statistical Science* 19:95–110.

Müller, Peter, Fernando Quintana, and Gary L Rosner. 2007. Semiparametric Bayesian Inference for Multilevel Repeated Measurement Data. *Biometrics* 63:280–289.

Mundlak, Yair. 1978. On the pooling of time series and cross section data. *Econometrica* 46:69–85.

Navarro, Daniel J, Thomas L Griffiths, Mark Steyvers, and Michael D Lee. 2006. Modeling individual differences using Dirichlet processes. *Journal of Mathematical Psychology* 50:101–122.

Neal, Radford M. 2000. Markov Chain Sampling Methods for Dirichlet Process Mixture Models. *Journal of Computational and Graphical Statistics* 9:249—-265.

Nerlove, Marc, Patrick Sevestre, and Pietro Balestra. 2008. Introduction. In *The Econometrics of Panel Data. Fundamentals and Recent Developments in Theory and Practice*, edited by Laszlo Matyas and Patrick Sevestre. Berlin: Springer, 3–22.

Neustadt, Ilja. 2010. Do Religious Beliefs Explain Preferences for Income Redistribution? Experimental Evidence. University of Zurich Socioeconomic Institute working paper 1009.

Nickell, Stephen. 1981. Biases in dynamic models with fixed effects. *Econometrica* 49:1417–1426.

Pang, Xun. 2010. Modeling Heterogeneity and Serial Correlation in Binary Time-Series Cross-sectional Data: A Bayesian Multilevel Model with AR(p) Errors. *Political Analysis* 18:470–498.

Pudney, Stephen. 2006. The dynamics of perception Modelling subjective well-being in a short panel. ISER Working Paper 2006-27.

Pudney, Stephen. 2008. The dynamics of perception: modelling subjective wellbeing in a short panel. *Journal of the Royal Statistical Society A* :21–40.

Rabe-Hesketh, Sophia and Skrondal. 2008. Generalized linear mixed effects models. In *Longitudinal Data Analysis: A Handbook of Modern Statistical Methods*, edited by Garret Fitzmaurice, Marie Davidian, Geert Verbeke, and Geert Molenberghs. Boca Raton: Chapman & Hall, 79–106.

Rehm, Philipp. 2011. Risk Inequality and the Polarized American Electorate. *British Journal of Political Science* 41:363–387.

Rehm, Philipp, Jacob S. Hacker, and Mark Schlesinger. 2012. Insecure Alliances: Risk, Inequality, and Support for the Welfare State. *American Political Science Review* 106:386–406.

Robert, Christan P. 2007. *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation*. New York: Springer.

Rossi, Peter E, Greg M Allenby, and Robert Mcculloch. 2005. *Bayesian Statistics and Marketing*. Chichester: Wiley.

Schervish, Mark J. 1995. *Theory of Statistics*. New York: Springer.

Scheve, Kenneth and David Stasavage. 2006. Religion and Preferences for Social Insurance. *Quarterly Journal of Political Science* 1:255–286.

Sethuraman, Jayaram. 1994. A Constructive Definition of Dirichlet Priors. *Statistica Sinica* 4:639–650.

Shayo, Moses. 2009. A Model of Social Identity with an Application to Political Economy: Nation, Class, and Redistribution. *American Political Science Review* 103:147–174.

Skrondal, Anders and Sophia Rabe-Hesketh. 2004. *Generalized latent variable modeling: Multilevel, longitudinal and structural equation models*. Boca Raton: Chapman & Hall.

Spiegelhalter, D J, A Thomas, N Best, and W R Gilks. 1997. *BUGS: Bayesian Inference Using Gibbs Sampling. Manual.* Cambridge: Medical Research Council Biostatistics Unit.

Spirling, Arthur and Kevin M Quinn. 2010. Identifying Intraparty Voting Blocs in the U.K. House of Commons. *Journal of the American Statistical Association* 105:447–457.

Stegmueller, Daniel. 2011. Apples and Oranges? The problem of equivalence in comparative research. *Political Analysis* 19:471–487.

Varin, Cristiano and Claudia Czado. 2010. A mixed autoregressive probit model for ordinal longitudinal data. *Biostatistics* 11:127–138.

Vella, F and Marno Verbeek. 1998. Whose wages do unions raise? A dynamic model of unionism and wage rate determination for young men. *Journal of Applied Econometrics* .

Vermunt, Jeroen. 2004. An EM algorithm for the estimation of parametric and nonparametric hierarchical nonlinear models. *Statistica Neerlandica* 58:220–233.

Vermunt, Jeroen, Bac Tran, and Jay Magidson. 2008. Latent Class Models in Longitudinal Research. In *Handbook of Longitudinal Research. Design, Measurement and Analysis*, edited by Scott Menard. Academic Press, 373–385.

Wawro, G. 2002. Estimating Dynamic Panel Data Models in Political Science. *Political Analysis* 10:25–48.

Winkelmann, Rainer. 2005. Subjective well-being and the family: Results from an ordered probit model with multiple random effects. *Empirical Economics* 30:749–761.

Wlezien, Christopher. 1995. The Public as Thermostat: Dynamics of Preferences for Spending. *American Journal of Political Science* 39:981–1000.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Wooldridge, Jeffrey M. 2005. Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity. *Journal of Applied Econometrics* 20:39–54.

APPENDICES

A. INITIAL OBSERVATIONS

To explicate the role of initial observations, rewrite the dynamic model

$$z_{it} = \phi z_{it-1} + \boldsymbol{x}_{it}\beta + \xi_i + \epsilon_{it}, t = 1, \ldots, T$$

in its explicit distributed lag representation by successive backward substitution (e.g., following Harris et al. 2008: 251):

$$z_{it} = \phi^t z_{i0} + \sum_{j=0}^{t-1} \phi^j \boldsymbol{x}_{it-j}\boldsymbol{\beta} + \frac{1-\phi^t}{1-\phi}\xi_i + \eta_{it} \tag{23}$$

with $\eta_{it} = \phi\eta_{it-1} + \epsilon_{it}$ with $\eta_{i0} = 0$.

This makes obvious that each observation of z_i can be expressed as the sum of several factors. The first part of equation (23), $\phi^t z_{i0}$ depends on the initial observation of the panel, while the second part depends on current and past covariate values. The third part $\frac{1-\phi^t}{1-\phi}\xi_i$ indicates proportional dependence on unobserved individual specific effects.

Direct estimation of (23) would require sufficiently large T and that ϕ^t decays sufficiently rapidly with t. Alternatively, one can specify an empirical approximation of z_{i0} (Pudney 2008: 27). Heckman's (1981b) approximation for $z_{i0}|\boldsymbol{x}_{it}, \xi_i$,

$$z_{i0} = \boldsymbol{\delta}'\boldsymbol{w}_i + \lambda\xi_i + \epsilon_{i0}, \tag{24}$$

as given in the main text, is obtained by first writing

$$z_{i0} = \boldsymbol{\delta}'\boldsymbol{w}_i + \eta_i \tag{25}$$

where $\boldsymbol{w}_i = (\boldsymbol{x}_{i0}, \boldsymbol{v}_i)$ is a vector of initial condition covariates comprised of covariate values at sample entry $\boldsymbol{x}_{i0}$ and additional background information $\boldsymbol{v}_i$. η_i is an individual error component at the initial condition. Next, decompose η_i into an individual specific (time-constant) random effect and a stochastic disturbance at $t = 0$. Instead of introducing a second individual random effect, Heckman employs the orthogonal projection

$$\eta_i = \lambda \xi_i + \epsilon_{i0} \tag{26}$$

which specifies η_i as resulting from random disturbance ϵ_{i0} and individual specific effect ξ_i. The random disturbance term at the initial condition ϵ_{i0} is now uncorrelated with ξ_i by design, and assumed uncorrelated with other errors, i.e. $\text{Cov}(\epsilon_{i0}, \epsilon_{it}) = 0, \; \forall t > 0$. The individual specific random effects ξ_i are allowed to have a different scaling in the initial conditions equations by including a scale factor λ. Substituting (26) into (25) yields the reduced form equation (24) for initial observations used in the main text.

B. DIRICHLET PROCESS

In this appendix I describe the Dirichlet process in more detail.[27] A Dirichlet process random effects model can be understood as a (countably) *infinite* mixture of points. Thus I start from specifying a *finite* mixture of points model for random effects and set up the Dirichlet process model from there by letting the number of points $K \to \infty$.

A finite nonparametric random effects prior Start by specifying some flexible distribution G for the random effects:

$$\xi_i \sim G(\boldsymbol{\phi}) \tag{27}$$

with hyperparameters $\boldsymbol{\phi}$. G can be approximated arbitrarily close by specifying a finite sum of K point masses and weights π_k,

$$G(\boldsymbol{\pi}, \boldsymbol{\zeta}) = \sum_{k=1}^{K} \pi_k \delta_{\zeta_k} \tag{28}$$

with $\sum_{k=1}^{K} \pi_k = 1$ and where δ_{ζ_k} is the Dirac delta function yielding a point mass at ζ_k. Here, $\boldsymbol{\phi} = (\boldsymbol{\zeta}, \boldsymbol{\pi})$ and random effects ξ_i are sampled from this distribution and are equal to one of the ζ_k.

In a Bayesian setup (e.g. Lo 1984), one has to specify priors for the weights, such as:

$$\zeta_k \sim G_0 \tag{29}$$

$$\boldsymbol{\pi} \sim \text{Dirichlet}(\boldsymbol{\alpha}) \tag{30}$$

where each of K discrete locations ζ_k are sampled from some base distribution G_0. The prior

[27] This section builds on the excellent presentation in Navarro et al. (2006).

over weights is a Dirichlet distribution of dimension K with parameters $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_K)$:

$$p(\boldsymbol{\pi}|\boldsymbol{\alpha}) = \mathcal{L}(\boldsymbol{\alpha})^{-1} \left(\prod_{k=1}^{K} \pi_k^{\alpha_k - 1} \right) \mathbf{1}(\boldsymbol{\pi}) \tag{31}$$

where $\mathbf{1}(\boldsymbol{\pi})$ is an indicator function equal to one if weights sum to one and zero otherwise. $\mathcal{L}$ is a normalizing function given by:[28]

$$\mathcal{L}(\boldsymbol{\alpha}) = \int \left(\prod_{k=1}^{K} \pi_k^{\alpha_k - 1} \right) \mathbf{1}(\boldsymbol{\pi})\, d\boldsymbol{\pi} = \frac{\prod_{k=1}^{K} \Gamma(\alpha_k)}{\Gamma\left(\sum_{k=1}^{K} \alpha_k\right)} \tag{32}$$

The Dirichlet prior for the weights $\boldsymbol{\pi}$ is taken to be symmetric, i.e. we use a parameter vector of length K with $(\alpha/K, \ldots, \alpha/K)$, thus ensuring that the sum of the parameter vector will always be α (e.g. Ishwaran and Zarepour 2002).

Moving to the infinite case Having specified a prior for the finite case, we elicit a prior specification for the infinite point mixture case by letting $K \to \infty$.

First, to make the clustering structure of the model explicit, define membership indicators s_i, which indicate to which subcluster the ith random effect is assigned. For a random effect of individual i the probability of belonging to subcluster k is given by the weight π_k, and thus

$$p(s_i = k|\boldsymbol{\pi}) = \pi_k. \tag{33}$$

Using membership indicators, the prior in (29)–(30) becomes:

$$\zeta_k \sim G_0 \tag{34}$$
$$\boldsymbol{\pi} \sim \text{Dirichlet}(\alpha/K) \tag{35}$$
$$s_i \sim \text{Multinomial}(\boldsymbol{\pi}) \tag{36}$$

where membership indicators are sampled from a multinomial with size one.

Second, we integrate out the subcluster weights $\boldsymbol{\pi}$ to get the conditional subcluster assignment probability when having already observed $N-1$ random effects assignments $S_{[i]} = \{s_1, \ldots, s_{N-1}\}$:

$$p(s_i = k|S_{[i]}, \alpha, K) = \int p(s_i = k|\boldsymbol{\pi}) p(\boldsymbol{\pi}|S_{[i]}, \alpha, K) d\boldsymbol{\pi} \tag{37}$$

To solve the integral, note that the first term of the integrand is π_k (cf. equation (33)). The

[28] See, e.g. Gill (2008a: 180). Γ is the gamma function, which is a generalization of the factorial function: for a non-negative integer n, $\Gamma(n) = (n-1)!$.

second term is the posterior probability

$$p(\boldsymbol{\pi}|S_{[i]}, \alpha, K) \propto p(S_{[i]}|\boldsymbol{\pi})p(\boldsymbol{\pi}|\alpha, K), \tag{38}$$

i.e. the product of a multinomial and Dirichlet distribution, which implies that the posterior distribution is also a Dirichlet (i.e. conjugacy of the resulting posterior).

Denote by $m_k = \#\{\xi_1 = \zeta_k\}$ the number of random effects assigned to subcluster k, and let $\boldsymbol{m} = (m_1, \ldots, m_K)$ be a 'member size' vector giving the number of individuals in each subcluster. The posterior probability $p(\boldsymbol{\pi}|\boldsymbol{s}_{/i}, \alpha, K)$ is distributed Dirichlet with parameter vector $\boldsymbol{s} + \alpha/K$. Thus

$$p(s_i = k|\boldsymbol{s}_{/i}, \alpha, K) \tag{39}$$

$$= \mathcal{L}(\boldsymbol{m} + \alpha/K)^{-1} \int \pi_k \left(\prod_l \pi_l^{m_l + \alpha/K - 1} \right) \mathbf{1}(\boldsymbol{\pi}) d\boldsymbol{\pi} \tag{40}$$

$$= \frac{\mathcal{L}(\boldsymbol{m} + \alpha/K + \mathbf{1}(k))}{\mathcal{L}(\boldsymbol{m} + \alpha/K)} \tag{41}$$

$$= \frac{m_k + \alpha/K}{N - 1 + \alpha} \tag{42}$$

where $\mathbf{1}(k)$ is an indicator vector (with length K) with a 1 at position k and zero otherwise.

Having integrated out the weights, consider now the limiting probability that random effect ξ_i gets assigned value(s) ζ_k of an existing subcluster k with $m_k \geq 1$:

$$p(s_i = k|S_{[i]}, \alpha) = \lim_{K\to\infty} \left(\frac{m_k + \alpha/K}{N - 1 + \alpha} \right) \tag{43}$$

$$= \frac{m_k}{N - 1 + \alpha} \tag{44}$$

Conversely, consider the limit probability that ξ_i gets assigned values from a new subcluster. Let $K_{/i}$ be the realized number of subclusters when $N - 1$ random effects have already been assigned. Denote by $\mathcal{S}$ the set of subclusters with $m_k = 0$ (i.e. the $K - K_{[i]}$ empty subclusters). The assignment probability for the ith random effect is then

$$p(s_i \in \mathcal{S}|S_{[i]}, \alpha) = \lim_{K\to\infty} \left(\sum_{l \in \mathcal{S}} \frac{m_l + \alpha/K}{N - 1 + \alpha} \right) \tag{45}$$

$$= \frac{\alpha}{N - 1 + \alpha} \lim_{K\to\infty} \left(\frac{K - K_{[i]}}{K} \right) \tag{46}$$

$$= \frac{\alpha}{N - 1 + \alpha} \tag{47}$$

Integrating out subcluster assignment indicator variables s_i yields the prior distribution for assigning a value to random effect ξ_i given that all other random effects $\boldsymbol{\xi}_{[i]}$ have already been assigned. This distribution is a mixture of the base distribution G_0 and the empirical distribution of $N-1$ previously assigned random effect values:

$$\xi_i|\boldsymbol{\xi}_{[i]}, \alpha, G_0 \sim \frac{\alpha}{N-1+\alpha} G_0 + \sum_{k=1}^{K_{[i]}} \frac{m_k}{N-1+\alpha} \delta_{\zeta_k}. \tag{48}$$

Drawing a sequence of random effects assignments from (48) yields a Polya urn scheme with parameters α and G_0 (Blackwell and MacQueen 1973). Using this scheme allows us to choose a prior for the random effects distribution G. We require that the marginal prior over parameters $(\zeta_1, \dots, \zeta_\infty)$ follows a Polya urn scheme. Blackwell and MacQueen (1973) show that the Dirichlet process does, and we can thus specify the Dirichlet process as nonparametric random effects prior:

$$\xi_i \sim G \tag{49}$$

$$G \sim DP(\alpha, G_0) \tag{50}$$

Dirichlet process The Dirichlet process is a stochastic process (a distribution over function spaces) whose sample paths (i.e. random functions draws) are probability measures with probability 1 (Ferguson 1973, 1974). Intuitively, it is a distribution over distributions, where each draw yields a Dirichlet distribution. More formally, let $(\Sigma, \mathcal{B})$ be a (measurable) space, and let G_0 be a random probability measure over it, and let α be a positive real number. A Dirichlet Process is a distribution G over $(\Sigma, \mathcal{B})$ such that for every (finite measurable) partition $(B_1, \dots, B_N)$:

$$(G(B_1), \dots, G(B_N)) \sim \text{Dirichlet}(\alpha G_0(B_1), \dots, \alpha G_0(B_N)). \tag{51}$$

G_0 can be interpreted as mean of the process, since for any measurable B, $E(G(B)) = G_0(B)$. The 'dispersion', 'strength' or 'prior mass' parameter α can be understood as inverse variance, since $V(G(B)) = G_0(B)/(\alpha+1)$, so that larger values of α imply a tighter concentration of the DP around G_0.

The posterior process for a drawing G from the DP and a subsequent random effect draw ξ_1 from G is a standard Dirichlet update (see Schervish 1995):

$$G|\xi_1 \sim DP\left(\alpha G_0 + \delta_{\xi_1}\right). \tag{52}$$

Iterating the updating yields

$$G|\xi_1, \dots, \xi_N \sim DP\left(\alpha G_0 + \sum_{i=1}^{N} \delta_{\xi_i}\right). \tag{53}$$

To see the connection to the infinite mixture model consider the predictive distribution for a new ξ_{N+1} given previous random effect realizations $\boldsymbol{\xi}$, with G marginalized out. For any $B \subset \Sigma$ we again get the Polya/Blackwell MacQueen (1973) urn scheme (cf. equation 48):

$$E(G(B)|\xi_1, \ldots, \xi_N) = \frac{\alpha G_0(B) + \sum_{i=1}^{N} \delta_{\xi_i}(B)}{\alpha + N} \tag{54}$$

$$\rightarrow \sum_{k=1}^{\infty} \pi_k \, \delta_{\zeta_k}(B) \tag{55}$$

with $\pi_k = \lim_{N\to\infty} m_k/N$, and where ζ_k represents one unique random effect value, and $m_k = \#\{\xi_i = \zeta_k\}$ in the sequence $(\xi_1, \ldots, \xi_N)$. A countably infinite mixture of the above form, which fulfils the definition of the Dirichlet Process, can be constructed by the stick-breaking random measure, as shown by Sethuraman (1994).

Stick breaking construction It is used to construct the infinite number of weights in (55). Let

$$v_k \sim \text{Beta}(1, \alpha), \quad k = 1, 2, \ldots \tag{56}$$

be an infinite sequence of beta distributed random variables. Set $\pi_1 = v_1$ and construct the remaining π_k via

$$\pi_k = v_k \prod_{l=1}^{k-1} (1 - v_l), \quad k = 2, 3, \ldots . \tag{57}$$

Let $\zeta_k \sim G_0$ and $G = \sum_{k=1}^{\infty} \pi_k \delta(\zeta_k)$; then $G \sim DP(\alpha, G_0)$. This constructive scheme implies that, just as the finite case in (28)–(30), G has now a clear definition as a random measure, since

$$\sum_{k=1}^{\infty} \pi_k = 1 \text{ wp } 1. \tag{58}$$

To see this note that

$$1 - \sum_{k=1}^{K} = 1 - v_1 - v_2(1 - v_1) - v_3(1 - v_2)(1 - v_1) - \cdots \tag{59}$$

$$= (1 - v_1)(1 - v_2 - v_3(1 - v_2) - \cdots) \tag{60}$$

$$= \prod_{k=1}^{K} (1 - v_k) \tag{61}$$

$$\rightarrow 0 \text{ wp } 1 \text{ as } K \rightarrow \infty. \tag{62}$$

Estimation via truncated Dirichlet process The stick breaking construction suggests an approximate sampling strategy for posterior DP inference. Choose a truncation value T for K, and set $v_T = 1$ to ensure that weights do sum to one. Then we have a finite representation of the infinite mixture of points:

$$G = \sum_{k=1}^{T} \pi_k \delta_{\zeta_k}, \tag{63}$$

where $\pi_k = 0$ for $k > T$. More details are given by Ishwaran and James (2001) and and Ishwaran and Zarepour (2002). This approximation yields good approximations even with low values for T, and is computationally tractable and can be implemented in available general purpose Bayesian inference packages such as JAGS, WinBUGS or PyMC. Discussions of other, more sophisticated sampling strategies (which require tailored code) are given in Escobar and West (1995), MacEachern and Müller (1998), Neal (2000), and Kyung et al. (2011).

In any 'real-life' political science application, one should check if the truncation threshold T was chosen large enough. A straightforward way is to sample from a model where T is set at twice the size, and investigate if the posterior samples of K – the sampled number of subclusters – are larger for this model. Figure 4 shows a histogram of the posterior distribution of K from just such a model run, where I set $T = 40$. It indicates that even with a higher truncation thresholds, the Dirichlet process never created more than 20 subclusters (the maximum sampled value of K is 17). Thus, the truncation level used in the main part of the paper is a good approximation.[29]

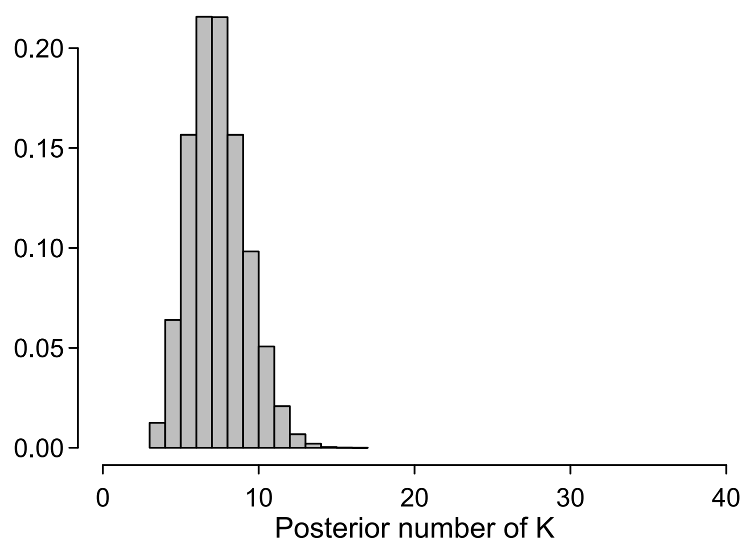

Figure 4: Posterior number of subclusters sampled from TDP(α, G_0, $T = 40$)

C. ELICITATION OF PARAMETERS OF Γ PRIOR FOR α

Kottas et al. (2005) derive an approximation of the mean and variance of the number of

[29] Furthermore, inspection of parameter estimates reveled no differences to a model with $T = 20$.

subclusters, which can be used to select semi-informative values for the Gamma prior of α. The expected number of subclusters given precision, α, and number of observations, N, is

$$E(k|\alpha, N) = \sum_{i=1}^{N} \frac{\alpha}{\alpha + i - 1} \approx \alpha \log\left(\frac{\alpha + N}{\alpha}\right) \tag{64}$$

with variance

$$\operatorname{Var}(k|\alpha, N) = \sum_{i=1}^{N} \frac{\alpha(i-1)}{(\alpha + i - 1)^2} \approx \alpha \left[\log\left(\frac{\alpha + N}{\alpha}\right) - 1\right]. \tag{65}$$

As a result of my Gamma prior specification $E(\alpha) = a_0/b_0$ and $\operatorname{Var}(\alpha) = a_0/b_0^2$. Some algebra yields the *a priori* expected mean and variance for the number of subclusters (cf. Kottas et al. 2005; Liu 1996: 916):

$$E(k) \approx \frac{a_0}{b_0} \log\left(1 + \frac{nb_0}{a_0}\right) \tag{66}$$

$$\operatorname{Var}(k) \approx \frac{a_0}{b_0} \log\left(1 + \frac{nb_0}{a_0}\right) - \frac{nb_0}{a_0} + \left[\log\left(1 + \frac{nb_0}{a_0}\right) - \frac{nb_0}{a_0 + nb_0}\right]^2 \frac{a_0}{b_0^2} \tag{67}$$

This expressions can be evaluated numerically to obtain reasonable values for a_0 and b_0 given ones prior expectations of the mean number of subclusters.[30]

D. INVERSE-GAMMA VARIANCE PRIORS

As mentioned in subsection 2.3 there are good reasons to prefer more informative priors for the random effect variance. In this section, I describe the specification (or 'elicitation') of two sets of hyperprior values.

Usually one specifies a prior for the inverse variance, or precision. The Gamma distribution is a popular choice (e.g. Gelman et al. 2004: 579). With given *a-priori* values for the expected mean m_0 and variance v_0 of the random effect precision σ_ξ^{-2}, hyperprior values for $\Gamma(a_0, b_0)$ are given by:[31]

$$a_0 = m_0^2/v_0 \tag{68}$$

$$b_0 = v_0/m_0 \tag{69}$$

Alternatively, when specifying a prior for the variance directly the inverse gamma distribution

[30] If researchers feel uncomfortable with choosing values based on expectations about K, they can either rely on priors suggested in the literature such as $\Gamma(1, 1)$ or $\Gamma(2, 2)$, which prevent very small and large values (Ishwaran and Zarepour 2000).

[31] I use the same notation for shape and scale of the Gamma distribution (a_0, b_0) as in subsection 3.2 purely for notational convenience.

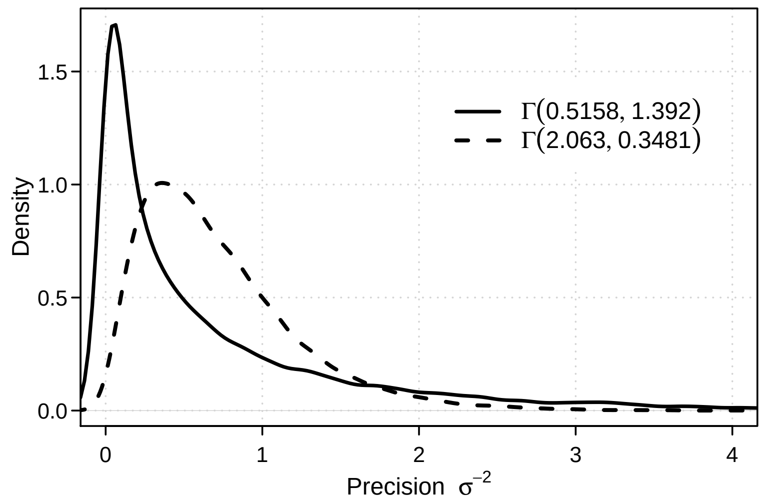

Figure 5: Distribution of variance prior precision under two Gamma prior specifications (based on 10,000 samples from prior distribution).

can be used. Here hyperprior values for $\Gamma^{-1}(a_0, b_0)$ are given by:

$$a_0 = (m_0^2 + 2v_0)/v_0 \tag{70}$$

$$b_0 = m_0(m_0^2 + v_0)/v_0 \tag{71}$$

A simple random effects ordered probit model fit using a laplace approximation to integrate out the random effects (ignoring the lagged dependent variable, and initial conditions) suggest a variance of the individual effects of ca. 1.392 or a precision of 0.7182. Thus, setting $m_0 = 0.7182$ I choose a two differently 'tight' v_0 values: $v_0 = \{1, 0.25\}$. This leads to hyperprior values of $a_0 = 0.5158$, $b_0 = 1.3924$, and $a_0 = 2.0632$, $b_0 = 0.3481$. The resulting prior distributions are illustrated in Figure 5 which plots 10,000 draws from the respective prior distributions.

Re-estimating my main model with these two more informative random effects variance prior choices leads to very similar estimated variances of 0.83 (sd=0.09) and 0.84 (sd=0.09), respectively. Coefficient estimates are virtually indistinguishable at two significant figures.

E. DP RANDOM EFFECTS ESTIMATES

Table 5 shows estimated parameters of the model with Dirichlet process random effects. α is the estimated dispersion parameter of the Dirichlet process; K represents the sampled value of the number of clusters at each MCMC step.

Table 5: Posterior summary for Hierarchical dynamic latent ordered probit model with Dirichlet process random effects.

Initial conditions		Mean	SD	95% HPD		Dynamics		Mean	SD	95% HPD	
δ_1	[Permanent income]	−0.403	0.130	−0.665	−0.155	β_1	[Permanent inc.]	−0.393	0.067	−0.526	−0.265
δ_2	[Transitory income]	−0.072	0.124	−0.312	0.172	β_2	[Transitory inc.]	−0.061	0.038	−0.135	0.016
δ_3	[Income share]	0.099	0.113	−0.118	0.326	β_3	[Income share]	−0.104	0.050	−0.203	−0.007
δ_4	[House value]	−0.402	0.241	−0.873	0.071	β_4	[House equity]	−0.110	0.052	−0.214	−0.011
δ_5	[House owner]	0.004	0.105	−0.204	0.209	β_5	[House owner]	−0.020	0.047	−0.111	0.074
δ_6	[HH size]	0.385	0.192	0.014	0.770	β_6	[HH size]	0.108	0.079	−0.050	0.260
δ_7	[N kids in HH]	−0.129	0.172	−0.471	0.204	β_7	[N kids in HH]	−0.032	0.066	−0.160	0.097
δ_8	[Union member]	0.171	0.101	−0.027	0.370	β_8	[Union member]	0.148	0.048	0.053	0.241
δ_9	[Age]	−0.720	0.126	−0.971	−0.479	β_9	[Age]	−0.336	0.056	−0.446	−0.225
δ_{10}	[Female]	0.485	0.113	0.265	0.710	β_{10}	[Female]	0.274	0.064	0.147	0.400
δ_{11}	[Divorced]	0.047	0.241	−0.427	0.509	β_{11}	[Divorced]	0.057	0.085	−0.100	0.231
δ_{12}	[Unemployed]	−0.408	0.221	−0.834	0.035	β_{12}	[Unemployed]	0.137	0.099	−0.058	0.331
δ_{13}	[Non-white]	0.134	0.293	−0.443	0.703	β_{13}	[Non-white]	0.482	0.171	0.152	0.820
δ_{14}	[Degree]	−0.499	0.173	−0.833	−0.156	β_{14}	[Degree]	−0.643	0.101	−0.845	−0.449
δ_{15}	[A-levels]	−0.546	0.171	−0.873	−0.203	β_{15}	[A-levels]	−0.469	0.102	−0.666	−0.267
δ_{16}	[O-levels]	−0.485	0.142	−0.760	−0.202	β_{16}	[O-levels]	−0.324	0.082	−0.480	−0.160
δ_{17}	[Parents: unskilled]	0.481	0.139	0.207	0.751	β_0	[Intercept]	0.405	0.073	0.264	0.551
δ_{18}	[Parents: skilled]	0.264	0.118	0.034	0.496	κ_2		0.811	0.018	0.777	0.844
δ_{19}	[Parents: white-collar]	0.324	0.127	0.077	0.572	ϕ		0.237	0.026	0.185	0.288
δ_{20}	[Parents: self-empl.]	0.160	0.131	−0.088	0.425	$\hat{r}$		−0.010	0.012	−0.034	0.012
δ_{21}	[London]	0.054	0.139	−0.224	0.320	α		1.476	0.508	0.583	2.502
λ		1.206	0.109	0.994	1.418	K		11.944	3.199	6.000	18.000
Deviance							25704				
DIC							26192				
Posterior predictive p-value							0.551				

Note: Based on 20.000 MCMC draws. Threshold τ_1 fixed at 0. Balanced panel. N1= 9044, N2=1292).