10-1-2012

Semiparametric Bayesian Modeling of Income Volatility Heterogeneity

Shane T. Jensen
University of Pennsylvania, stjensen@wharton.upenn.edu

Stephen H. Shore
Georgia State University, sshore@gsu.edu

Semiparametric Bayesian Modeling of Income Volatility Heterogeneity

Abstract
Research on income risk typically treats its proxy—income volatility, the expected magnitude of income changes—as if it were unchanged for an individual over time, the same for everyone at a point in time, or both. In reality, income risk evolves over time, and some people face more of it than others. To model heterogeneity and dynamics in (unobserved) income volatility, we develop a novel semiparametric Bayesian stochastic volatility model. Our Markovian hierarchical Dirichlet process (MHDP) prior augments the recently developed hierarchical Dirichlet process (HDP) prior to accommodate the serial dependence of panel data. We document dynamics and substantial heterogeneity in income volatility.

Keywords
hierarchical Dirichlet process, income volatility, state-space models

Disciplines
Demography, Population, and Ecology | Family, Life Course, and Society | Social and Behavioral Sciences | Sociology

Comments
Jensen, Shane T., and Stephen H. Shore. 2012. "Semiparametric Bayesian Modeling of Income Volatility Heterogeneity." PARC Working Paper Series, WPS 12-03.

Semiparametric Bayesian Modeling of Income Volatility Heterogeneity

Shane T. Jensen and Stephen H. Shore

October 2012

BWP2012-05
Boettner Center Working Paper
Boettner Center for Pensions and Retirement Research
The Wharton School, University of Pennsylvania
3620 Locust Walk, 3000 SH-DH
Philadelphia, PA 19104-6302
Tel: 215.898.7620 Fax: 215.573.3418
Email: prc@wharton.upenn.edu
http://www.pensionresearchcouncil.org

The project described was supported by the National Institute on Aging, P30 AG-012836-18, the Boettner Center for Pensions and Retirement Security, and National Institutes of Health – National Institute of Child Health and Development Population Research Infrastructure Program R24 HD-044964-9, all at the University of Pennsylvania. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Aging or the National Institutes of Health. All findings, interpretations, and conclusions of this paper also do not represent the views of the Wharton School or the Boettner Center for Pensions and Retirement Research.

Semiparametric Bayesian Modeling of Income Volatility Heterogeneity

Abstract

Research on income risk typically treats its proxy—income volatility, the expected magnitude of income changes—as if it were unchanged for an individual over time, the same for everyone at a point in time, or both. In reality, income risk evolves over time, and some people face more of it than others. To model heterogeneity and dynamics in (unobserved) income volatility, we develop a novel semiparametric Bayesian stochastic volatility model. Our Markovian hierarchical Dirichlet process (MHDP) prior augments the recently developed hierarchical Dirichlet process (HDP) prior to accommodate the serial dependence of panel data. We document dynamics and substantial heterogeneity in income volatility.

Keywords: hierarchical Dirichlet process; income volatility; state-space models.

Shane T. Jensen
The Wharton School, University of Pennsylvania
Department of Statistics
Philadelphia, PA 19104
stjensen@wharton.upenn.edu

Stephen H. Shore
The Robinson College of Business
Georgia State University
Department of Risk Management and Insurance
Atlanta, GA 30303
sshore@gsu.edu

Semiparametric Bayesian Modeling of Income Volatility Heterogeneity

Shane T. JENSEN and Stephen H. SHORE

Research on income risk typically treats its proxy—income volatility, the expected magnitude of income changes—as if it were unchanged for an individual over time, the same for everyone at a point in time, or both. In reality, income risk evolves over time, and some people face more of it than others. To model heterogeneity and dynamics in (unobserved) income volatility, we develop a novel semiparametric Bayesian stochastic volatility model. Our Markovian hierarchical Dirichlet process (MHDP) prior augments the recently developed hierarchical Dirichlet process (HDP) prior to accommodate the serial dependence of panel data. We document dynamics and substantial heterogeneity in income volatility.

KEY WORDS: Hierarchical Dirichlet process; Income volatility; State-space models.

1. INTRODUCTION AND MOTIVATION

Income dynamics—how peoples' incomes evolve over time—are of great interest to economists. One reason for this is that income volatility—the variance of income changes—is frequently used as a proxy for income risk. To the degree that people are risk-averse, such risk may carry substantial welfare costs. Many social programs (e.g., unemployment insurance, progressive income taxes) are designed to mitigate such costs.

Attempts to measure income volatility typically assume that income volatility is the same for everyone, that it does not change from year to year, or both. When volatility is allowed to vary across time or individuals, such differences are tied to covariates (e.g., comparing the income volatility of high- and low-education individuals).[1] In this article we develop a new method to identify flexibly latent differences in income volatility across individuals and over time. Such heterogeneity has a dramatic impact on our interpretation of income volatility. To the degree that volatility is chosen, we would expect individuals to take on more of it when they are more risk-tolerant or have better risk-sharing opportunities. As a result, average measures of income volatility (coupled with average measures of risk aversion) may dramatically overstate the welfare cost of income risk. For example, Jensen and Shore (2009) used the methods outlined here to show that the increase in average income volatility over time can be attributed to a minority of individuals with high income volatility; these individuals are more likely to self-identify as risk-tolerant.

We examine self-reported labor incomes from the core sample of the Panel Study of Income Dynamics (PSID), designed as a nationally representative panel of U.S. households. Data were collected annually from 1968 to 1997 and biennially to 2005. Among the 3041 male household heads that we study, the average number of income observations per individual was 17. This rich panel allows us to model income dynamics for each individual and to compare those dynamics across individuals.

We build our current work on a standard model for income dynamics that decomposes changes in log income into predictable changes (with covariates), permanent shocks, and transitory shocks. The rate at which permanent and transitory shocks enter into and exit from income are governed by (homogeneous) parameters that we estimate. The chief objects of interest are permanent and transitory volatility, the variance of permanent and transitory income shocks. We develop a stochastic volatility model in which volatility parameters differ across individuals and evolve over time.

Conditional on volatility parameters, we decompose income changes in a dynamic linear model, deconvolving the permanent and transitory components of the income process for each individual. We estimate this dynamic linear model using a stochastic extension (Carter and Kohn 1994) of the standard Kalman filter (Kalman 1960), nested within our full model outlined below.

The main challenge in modeling the entire income volatility distribution is a lack of a priori knowledge of the appropriate functional form for this distribution. Absent such knowledge, we pursue a nonparametric strategy when modeling income volatility. Specifically, we build on the popular Bayesian nonparametric approach of specifying a completely unknown distribution for income volatility, with a Dirichlet process (DP) prior on this unknown distribution. DP priors have been used extensively for the estimation of unknown distributions, as reviewed by Muller and Quintana (2004). This prior induces a discreteness on the posterior distribution that has often been used to cluster observations or latent variables (see, e.g., Medvedovic and Sivaganesan 2002; Jensen and Liu 2008; Quintana et al. 2008). In our application, income volatility can take one of a discrete number of values, with the number of such values and the values themselves determined by the data.

A standard DP-based model does not account for the *grouped* nature of our labor income data. For each individual in our dataset, we have multiple observations of reported income over an individual-specific number of years. An individual's volatility value in year t is likely to be more similar to his or her value

Shane T. Jensen is Associate Professor, Department of Statistics, The Wharton School, University of Pennsylvania, Philadelphia, PA 19104 (E-mail: *stjensen@wharton.upenn.edu*). Stephen H. Shore is Associate Professor, Department of Risk Management and Insurance, J. Mack Robinson College of Business, Georgia State University, Atlanta, GA 30303 (E-mail: *sshore@gsu.edu*).

[1] There are a few exceptions. Meghir and Windemejier (1999), Banks, Blundell, and Grugiavini (2001), and Meghir and Pistaferri (2004) considered an autoregressive conditional heteroscedasticity (ARCH) process for volatility, and Browning, Alvarez, and Ejrnaes (2010) allowed for heterogeneity in the degree to which shocks are autoregressive.

Journal of the American Statistical Association
December 2011, Vol. 106, No. 496, Applications and Case Studies
DOI: 10.1198/jasa.2011.ap09283

in another year than to a value of another individual. Teh et al. (2006) recently addressed the situation of grouped data by introducing the hierarchical DP (HDP). This hierarchical extension of the standard DP-based model induces clustering of variables both within groups and between groups in the population. The HDP is a natural framework for a model-based approach to our analysis. Each group is an individual person, and so this flexible model allows us to share information on income volatility both within an individual (across time) and between individuals (across the population).

A standard HDP-based model does not account for the *panel* nature of our data. Each group (each individual's data) is ordered by year. An individual's volatility in year t is likely to be more similar to his or her volatility in year $t-1$ than in some other year distant from year t. We introduce a novel extension of the HDP framework that imposes a Markovian time dependence between volatilities within each individual. Volatility may remain unchanged from one year to the next. If volatility changes, it may change to a value held by that individual in another year or to a value not elsewhere present for that individual; if it changes to a value not elsewhere present for that individual, it may change to a value held by other individuals in the data or to a new value not seen elsewhere in the sample.

We combine our novel Markovian HDP (MHDP) model for volatility with a dynamic linear model for permanent versus transitory shocks. This semiparametric Bayesian model estimates the evolution of income dynamics over time across the entire population of individuals in our study. Our full model is implemented with a Gibbs sampler to estimate the full posterior distribution of each unknown parameter in our model.

There is an extensive literature on stochastic volatility models. The simplest of these, autoregressive conditional heteroscedasticity (ARCH) models, allow the current volatility to also be a parametric function of previous shocks (Engle 1982) and have been applied to income dynamics in Meghir and Windemejier (1999), Banks, Blundell, and Grugiavini (2001), and Meghir and Pistaferri (2004). Generalized (GARCH) versions of these models also allow the current volatility to be parametric functions of both previous shocks and previous volatility parameters (Bollerslev 1986). Compared with these simpler alternatives, our MHDP model allows flexibility in the shape of the cross-sectional volatility distribution and its evolution. Such flexibility could in principle be obtained with a sufficiently rich GARCH structure, but at considerable cost in terms of parsimony and ease of computation. In contrast, the MHDP model inherits from HDP models the intuitive estimation method and structure for sharing information across years and between individuals. As such, this methodological contribution allows us to tractably accommodate flexibility in volatility heterogeneity and dynamics.

The MHDP model is particularly appealing in the analysis of the volatility of individual income processes. For this application, the key feature of the MHDP is its posterior discreteness. Unlike an ARCH model in which volatility is constantly changing, volatility in an MHDP setting is constant for a period and then jumps to a different value. Such discontinuous changes in volatility can reflect discrete life events (career changes, beginning care of an elderly parent who falls ill). Having said that, findings on ARCH-based income dynamics (Meghir and Pistaferri 2004) mirror our own: they also reject the hypothesis that income volatility is the same for everyone and that it remains unchanged over time for an individual.

We estimate our model on data from the PSID. We find that the marginal distribution of volatility parameters is strongly positively skewed. Most individuals have volatility parameters that conform to our expectations (e.g., shocks with annual standard deviations of 15 percent), but a small minority of individuals have enormous volatility parameters (with standard deviations of 100 percent or more). Although volatility parameters are highly persistent, we can strongly reject the hypothesis that they are constant over time.

2. MODEL AND IMPLEMENTATION

Data are taken from the core sample of the PSID. The PSID tracked families annually from 1968 to 1997 and has done so in odd-numbered years thereafter. Our outcome variable is *excess* log income, the residual from a regression to predict the natural log of self-reported labor income with covariates (e.g., age, education). Details on sampling restrictions, covariates used to obtain excess log income, missing data, and outliers are presented in Section 3.1.

2.1 Income Process

Our model is based on a standard process for excess log income for individual i at time t (Carroll and Samwick 1997; Meghir and Pistaferri 2004). In this article, time t represents calendar time, not the age of individual i. Excess log income $y_{i,t}$ is modeled as the sum of permanent income, transitory income, and error $e_{i,t}$,

$$y_{i,t} = \underbrace{\sum_{k=1}^{t-3} \omega_{i,k} + \sum_{k=t-2}^{t} \phi_{\omega,t-k} \cdot \omega_{i,k}}_{\text{Permanent income}} + \underbrace{\sum_{k=t-2}^{t} \phi_{\varepsilon,t-k} \cdot \varepsilon_{i,k}}_{\text{Transitory income}} + e_{i,t}. \quad (1)$$

Permanent income is the weighted sum of past permanent shocks $\omega_{i,k}$ to income. Transitory income is the weighted sum of recent transitory shocks $\varepsilon_{i,k}$ to income. Although we use the word "shock" for parsimony, these innovations to income may be predictable to the individual, even if they look like shocks in the data.[2]

In our model, permanent shocks come into effect over three periods, and transitory shocks fade completely after three periods,[3] giving us three permanent weight parameters $(\phi_{\omega,0}, \phi_{\omega,1}, \phi_{\omega,2})$ and three transitory weight parameters $(\phi_{\varepsilon,0}, \phi_{\varepsilon,1}, \phi_{\varepsilon,2})$. We refer to these weights $\boldsymbol{\phi}$ collectively as the

[2] Furthermore, transitory shocks are observationally equivalent to measurement error when $\phi_{\varepsilon,t-k} = 0$ for $t-k > 0$. Whether measurement error or a transitory shock, income will change temporarily and then revert to its previous level. Our results on heterogeneity in transitory volatility are equivalent to documenting heterogeneity in the degree of measurement error.

[3] The choice to allow shocks to enter in and fade out over at most 3 years is motivated by the work of Abowd and Card (1989), who documented that income changes are not autocorrelated at lags greater than 2 years. We get very similar results using two, four, or five periods.

income process parameters,[4] which are shared globally by all individuals to make their estimation computationally tractable. We posit flat prior distributions for each weight parameter [i.e., $p(\phi) \propto 1$]; however, to give meaning to the magnitude of our transitory shocks, we normalize the weights placed on transitory shocks to sum to 1 ($\sum_k \phi_{\varepsilon,k} = 1$).

The permanent shock, transitory shock, and error term are assumed to be normally distributed as well as conditionally independent of one another over time and across individuals (given values of the other parameters). We examine this normality assumption at length in Section 4.1. The permanent shocks $\omega_{i,t}$ have mean 0 and *permanent variance* $\sigma^2_{\omega,i,t} \equiv \mathrm{E}[\omega^2_{i,t}]$, and the transitory shocks $\varepsilon_{i,t}$ have mean 0 and *transitory variance* $\sigma^2_{\varepsilon,i,t} \equiv \mathrm{E}[\varepsilon^2_{i,t}]$:

$$\begin{pmatrix} \omega_{i,t} \\ \varepsilon_{i,t} \end{pmatrix} \sim \mathrm{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{\omega,i,t} & 0 \\ 0 & \sigma^2_{\varepsilon,i,t} \end{bmatrix} \right). \tag{2}$$

We refer to $\sigma^2_{i,t} \equiv (\sigma^2_{\varepsilon,i,t}, \sigma^2_{\omega,i,t})$ jointly as the volatility parameters, which are the main objects of interest in this study. Subscripts for i and t indicate that volatility parameters may differ across individuals and over time, as discussed in Sections 2.3–2.4. This accommodates not just an evolving distribution of volatility parameters, but also systematic changes over the life cycle of volatility parameters, as documented by Shin and Solon (2008) and others. Finally, we have a homoscedastic noise shock $e_{i,t}$ with mean 0 and *residual variance* $\gamma^2 \equiv \mathrm{E}[e^2_{i,t}]$. For this residual variance parameter, we use a noninformative prior distribution $p(\gamma^2) \propto (\gamma^2)^{-1}$.

This parsimonious framework for income dynamics is not the only option. In particular, our approach rules out an autoregressive structure in income levels (Gottschalk and Moffitt 2002) and does not allow for heterogeneous rates of income growth (Baker 1997; Baker and Solon 2003). Abowd and Card (1989) noted an absence of autocorrelation in income changes, and this is frequently used to defend models of the class that we use. However, Baker (1997) argued that the Abowd and Card (1989) result merely reflects sample and power problems. Our choice of income process is more practical than ideological; both of the alternatives that we mention here require that we keep track of and model additional state variables (i.e., person-specific long-term income levels or growth rates), which would be computationally taxing. Because our model is estimated from relatively high-frequency income changes, the trends and mean reversion in alternative models should have little impact on our estimated parameters.

2.2 Formulation of Income Process as a State-Space Model

We first present our methodology for estimating the income process model of Section 2.1, assuming that the income volatility parameters $\sigma^2_{i,t}$ are known. We then generalize our model to allow for unknown volatilities in Sections 2.3–2.4. We reformulate our income process model (1) as

$$\begin{aligned} y_{i,t} &= \mathbf{X}_{i,t} \cdot \boldsymbol{\beta} + e_{i,t}, \quad \text{where} \\ \mathbf{X}_{i,t} &= \mathbf{A} \cdot \mathbf{X}_{i,t-1} + \mathbf{Q}_{i,t}. \end{aligned} \tag{3}$$

The vector $\boldsymbol{\beta}$ collects our income process weights,

$$\boldsymbol{\beta} = (1, \phi_{\omega,2}, \phi_{\omega,1}, \phi_{\omega,0}, \phi_{\varepsilon,2}, \phi_{\varepsilon,1}, \phi_{\varepsilon,0})',$$

and the vectors $\mathbf{Q}_{i,t}$ and $\mathbf{X}_{i,t}$ contain the latent shocks,

$$\mathbf{X}_{i,t} = \left(\sum_{s=1}^{t-3} \omega_{i,s}, \omega_{i,t-2}, \omega_{i,t-1}, \omega_{i,t}, \varepsilon_{i,t-2}, \varepsilon_{i,t-1}, \varepsilon_{i,t} \right)$$

and

$$\mathbf{Q}_{i,t} = (0, 0, 0, \omega_{i,t}, 0, 0, \varepsilon_{i,t}),$$

where $\omega_{i,t}$ and $\varepsilon_{i,t}$ are drawn from (2). The error term $e_{i,t}$ is normally distributed with mean 0 and variance γ^2. $\mathbf{A}$ is a matrix of constants that does not have to be estimated ($\mathbf{A}$ encodes the transition that takes $\omega_{i,t}$ and $\varepsilon_{i,t}$ when $t = 1$ to $\omega_{i,t-1}$ and $\varepsilon_{i,t-1}$ when $t = 2$, etc.). The formulation (3) can be recognized as a *state-space model* where the permanent and transitory shocks are latent states collected in $\mathbf{X}_{i,t}$, which evolve over time. The state-space formulation has been used extensively in stochastic volatility models (Durbin and Koopman 2001). However, this formulation is predicated on knowing the values of our income volatility parameters $\sigma^2_{i,t}$, which we address with a flexible semiparametric approach in Sections 2.3–2.4.

2.3 Modeling Volatility With a Hierarchical Dirichlet Process

Aside from a few exceptions using an ARCH structure (Meghir and Windemejier 1999; Banks, Blundell, and Grugiavini 2001; Meghir and Pistaferri 2004), previous research modeling income dynamics has assumed that all individuals with the same demographics have the same volatility parameters, $\sigma^2 \equiv (\sigma^2_\omega, \sigma^2_\varepsilon)$. The primary focus of this article is on allowing heterogeneity of these volatility parameters, $\sigma^2_{i,t} \equiv (\sigma^2_{\varepsilon,i,t}, \sigma^2_{\omega,i,t})$, both across individuals and over time within an individual.[5] These individual volatility parameters are not actually known, as was assumed in Section 2.2, and so we need to formulate a probability model for the distribution of volatilities $\sigma^2_{i,t}$ in the population. Without an a priori view of the correct functional form for the distribution of $\sigma^2_{i,t}$, we take a nonparametric modeling approach. We can model $\sigma^2_{i,t}$ as iid draws from a completely unknown distribution,

$$\sigma^2_{i,t} \sim \mathrm{G}(\cdot). \tag{4}$$

This unknown distribution $\mathrm{G}(\cdot)$ represents the common structure between the different volatility parameters in the population. A popular Bayesian approach to nonparametric problems is to give the unknown distribution $\mathrm{G}(\cdot)$ a DP prior, $\mathcal{D}(\alpha_D \cdot \mathrm{H})$, where H is a finite (nonnegative) probability measure (Ferguson

[4] As an example of our notation, $\phi_{\omega,2}$ denotes the weight placed on a permanent shock from two periods ago $\omega_{i,t-2}$ in current excess log income, and $\phi_{\varepsilon,2}$ denotes the weight placed on a transitory shock from two periods ago $\varepsilon_{i,t-2}$ in current excess log income. Carroll and Samwick (1997) assumed $\phi_{\omega,k} = \phi_{\varepsilon,k} = 0$ for $k > 0$, although they acknowledged that this assumption is unrealistic and designed an estimation strategy that is robust to this restriction, but did not estimate ϕ_k. Meghir and Pistaferri (2004) and Blundell, Pistaferri, and Preston (2008) assumed $\phi_{\omega,k} = 0$ for $k > 0$ but did not assume $\phi_{\varepsilon,k} = 0$.

[5] An alternative approach was taken by Browning, Alvarez, and Ejrnaes (2010), who allowed heterogeneity in parameters similar to ϕ and were able to reject the hypothesis that these values are the same for everyone. We abstract from ϕ heterogeneity to maintain tractability while considering heterogeneity in volatility.

1974).[6] A reasonable choice of H for a population of variance parameters is a χ^2_ν distribution with small degrees of freedom parameter ν. The parameter α_D is a weighting factor that characterizes our prior confidence in the shape of G. A well-known consequence of this model is that the posterior distribution of $G(\cdot)$ is discretized. As explained by Ferguson (1974), if $\theta_1, \ldots, \theta_n$ are n iid observations from the probability function G whose prior distribution is the DP $\mathcal{D}(\alpha_D \cdot H)$, then

$$G(\cdot)|\theta_1, \ldots, \theta_n \sim \mathcal{D}\left(\alpha_D \cdot H + \sum_{j=1}^{n} \boldsymbol{\delta}_{\theta_j}\right). \tag{5}$$

Thus the posterior mean of $G(\cdot)$, or the predictive distribution of a new observation, is proportional to $\alpha_D \cdot H + \sum_{j=1}^{n} \boldsymbol{\delta}_{\theta_j}$. The point mass components allow for the clustering of similar variables. The parameter α_D (D for "Dirichlet") implicitly controls the expected number of these clusters in the population.

Although the standard DP prior is a conventional choice for modeling of an unknown distribution, it does not respect several key features of our data. First, our data are *grouped*; each group is an individual tracked across several years. If individuals differ systematically in their volatility parameters $\sigma^2_{i,t}$, then the assumption of iid volatility values is not appropriate. Second, our data are *ordered*. An individual's volatility in year t is likely to be more similar to his or her volatility in year $t-1$ than in some other year distant from year t.

We can address the grouped structure of the data using an HDP prior, as outlined by Teh et al. (2006). In an HDP, $\sigma^2_{i,t}$ values are taken as iid draws from a group- (in this case, individual-) specific unknown distribution,

$$\sigma^2_{i,t} \sim G_i(\cdot). \tag{6}$$

This unknown distribution $G_i(\cdot)$ represents the common structure between the different volatility parameters across time within each individual i. Information is then shared between individuals by imposing another level of the prior model in which all the prior measures G_i also share a common DP prior distribution,

$$G_i \sim \mathcal{D}(\alpha_H \cdot G_0), \tag{7}$$

where the distribution G_0 itself has a DP prior distribution

$$G_0 \sim \mathcal{D}(\alpha_D \cdot H), \tag{8}$$

where H is a base measure for the population of variance parameters. A reasonable choice for H is a χ^2_ν distribution with small degrees of freedom parameter ν. The combination of prior distributions (7)–(8) gives an overall prior structure that allows clustering of volatility parameters across years within an individual as well as across individuals within the population. This clustering is influenced by two parameters. α_H (H for "hierarchical") implicitly controls the expected number of volatility clusters per group (individual), whereas α_D implicitly controls the expected number of volatility clusters in the population.

[6] Hirano (2002) also used such a DP prior in a model of income dynamics. Our motivation and use of this framework is completely different, however. Hirano used a DP prior to accomodate flexibility in the shape of the distribution of shocks conditional on volatility; in contrast, we use it to accommodate flexibility in the distribution of volatility.

A different nonparametric approach for grouped data is the nested DP given by Rodriguez, Dunson, and Gelfand (2008), where a clustering is imposed on entire groups in addition to being imposed on observations within each group. We prefer the HDP, in which volatility values $\sigma^2_{i,t}$ are shared individually between individuals and over time within an individual.

However, our formulation up to this point does not address the ordered panel structure of our data. In the next section, we introduce a novel Markovian extension of the HDP prior that allows $\sigma^2_{i,t}$ to remain unchanged from the previous period ($\sigma^2_{i,t} = \sigma^2_{i,t-1}$).

2.4 Markovian Hierarchical Dirichlet Process

The HDP outlined in (6)–(8) does not acknowledge any ordering of an individual's volatilities, but it is reasonable to assert that an individual's volatility in year t is likely to be more similar to his or her volatility in year $t-1$ than in some other year farther away from year t. We introduce an MHDP model that extends the HDP framework to accommodate panel data, introducing a time dependence between volatilities within each individual.

We start from the HDP model specified earlier,

$$G_0 \sim \mathcal{D}(\alpha_D \cdot H),$$
$$G_i \sim \mathcal{D}(\alpha_H \cdot G_0).$$

Now, instead of generating $\sigma^2_{i,t}$ from individual-specific G_i (ignoring the time ordering), we generate $\sigma^2_{i,t}$ conditional on that individual's volatility history up to time point t. Specifically, we have a Bernoulli variable, $Z_{i,t}$, at each time point, where $Z_{i,t} = 1$ indicates that $\sigma^2_{i,t}$ is the same as previous volatility value $\sigma^2_{i,t-1}$. We sample these Bernoulli variables $Z_{i,t}$ with probabilities

$$\begin{aligned} p(Z_{i,t} = 1|\mathbf{Z}_{i,1:t-1}, \alpha_M) &\propto N^\star_{it}, \\ p(Z_{i,t} = 0|\mathbf{Z}_{i,1:t-1}, \alpha_M) &\propto \alpha_M, \end{aligned} \tag{9}$$

where $N^\star_{it}$ is now the number of consecutive preceding years where $Z_{i,k} = 1$ (i.e., where $\sigma^2_{i,k} = \sigma^2_{i,t-1}$). α_M (M for "Markovian") is the prior weight on the choice to look beyond the previous year's value. If $Z_{i,t} = 1$, we set $\sigma^2_{i,t} = \sigma^2_{i,t-1}$. If $Z_{i,t} = 0$, we sample $\sigma^2_{i,t} \sim G_i$.

This modification of our model for generating $\sigma^2_{i,t}$ induces a Markovian dependency within the hierarchical Dirichlet process. The resulting process is still discrete and imposes a clustering on the volatility parameters $\boldsymbol{\sigma}^2$. However, the added Markovian dependency ensures not only that particular volatility values are likely to be conserved within an individual's history (as would be with the HDP), but also that contiguous years within an individual are likely to share the same volatility value.

Building time ordering into a DP model has substantial precedence. Fox et al. (2007) extended a HDP-HMM model for speaker diarization to give extra preference to self-transitions between neighboring time points. Other recent work has explicitly modeled the evolution of DP clusters over time (Wang and McCallum 2006; Ahmed and Xing 2008) and dependence as a function of covariates (MacEachern 1999; Iorio et al. 2004). Griffin and Steel (2006) also developed a DP construction conditional on a covariate (e.g., time), and Xue, Dunson, and Carin (2007) developed a dependent DP model for multitask learning.

Zhu, Ghahramani, and Lafferty (2005) and Blei and Frazier (2010) discussed different weighting functions for ordered data within a DP model. Neither of these approaches involves a hierarchical DP or our specific form (9) for time dependence. Our weighting scheme based on the current streak, $N^{\star}_{it}$, of the previous value is somewhat ad hoc, but reflects our desire to encourage extra stickiness toward last year's value when considering the current year's value. We experimented with other dependence structures and found generally similar results.

In the context of other stochastic volatility models, such as the ARCH (Engle 1982) or generalized ARCH (GARCH) (Bollerslev 1986) frameworks, our MHDP model maintains the relative simplicity (and ease of estimation) of the HDP framework while allowing an extremely flexible (and time-varying) shape of the cross-sectional volatility distribution.

This Markovian approach provides an alternative to the ARCH volatility dynamics used by Meghir and Pistaferri (2004). One key disadvantage of our approach is that when volatility changes, the previous level is irrelevant; presumably high volatility values are likely to be followed by high values even when volatility changes. Two advantages of our approach are that we provide a tractable way to capture a flexible cross-sectional distribution for volatility and allow for regime changes in volatility that may be helpful in modeling career-switching.

2.5 Variable Weight Parameters

The MHDP is influenced by three weight parameters: α_M, which implicitly controls the probability that volatility changes from one year to the next; α_H, which implicitly controls the expected number of volatility clusters per individual; and α_D, which implicitly controls the expected number of volatility clusters in the population. Thus far, we have presented these prior weight parameters $\boldsymbol{\alpha} = (\alpha_M, \alpha_H, \alpha_D)$ as fixed and known values.

Similar to the approach of Escobar and West (1995), we now allow these weight parameters to vary with their own prior distributions. Specifically, $\alpha_k \sim \text{Gamma}(a, b)$ for each $k \in (M, H, D)$. We set $a = b = 1/2$, which gives us prior expectation $\text{E}(\alpha_k) = 1$ and prior variance $\text{Var}(\alpha_k) = 2$ for each weight parameter. We also examined other values for these hyperparameters a and b and found that they did not affect our results presented in Section 3.

Our complete model contains several sets of unknown parameters, including the global income process parameters $\boldsymbol{\phi}$, noise parameter γ^2, and weight parameters $\boldsymbol{\alpha}$ shared across all individuals and all years, as well as the permanent and transitory latent shocks, $\boldsymbol{\omega}$ and $\boldsymbol{\epsilon}$, and their corresponding volatility parameters, $\boldsymbol{\sigma}^2 = (\sigma^2_{\omega}, \sigma^2_{\varepsilon})$ that vary for each individual in each year. The full posterior distribution of these unknown parameters can be constructed from the levels of our hierarchical model,

$$\begin{aligned} &p(\boldsymbol{\phi}, \gamma^2, \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\varepsilon}, \boldsymbol{\sigma}^2 | \mathbf{y}) \\ &\quad \propto p(\mathbf{y} | \boldsymbol{\phi}, \gamma^2, \boldsymbol{\omega}, \boldsymbol{\varepsilon}) \cdot p(\boldsymbol{\omega}, \boldsymbol{\varepsilon} | \boldsymbol{\sigma}^2) \cdot p(\boldsymbol{\sigma}^2 | \boldsymbol{\alpha}) \\ &\qquad \cdot p(\boldsymbol{\alpha}) \cdot p(\boldsymbol{\phi}) \cdot p(\gamma^2), \end{aligned} \tag{10}$$

where $p(\mathbf{y}|\boldsymbol{\phi}, \gamma^2, \boldsymbol{\omega}, \boldsymbol{\varepsilon})$ and $p(\boldsymbol{\omega}, \boldsymbol{\varepsilon}|\boldsymbol{\sigma}^2)$ are determined by our income process state-space model (Section 2.2), with the noninformative prior distributions $p(\boldsymbol{\phi})$ and $p(\gamma^2)$ outlined in Section 2.1. The prior distribution $p(\boldsymbol{\sigma}^2|\boldsymbol{\alpha})$ is determined by the MHDP in Sections 2.3–2.4, and $p(\boldsymbol{\alpha})$ is given in Section 2.5.

2.6 Gibbs Sampling Implementation

We estimate the full posterior distribution (10) using Gibbs sampling (Geman and Geman 1984), with sets of unknown parameters sampled conditional on the current values of all other parameters. For our model, we sample our unknown parameters iteratively as follows:

1. Sample income process shocks $(\boldsymbol{\omega}, \boldsymbol{\varepsilon})$ conditional on $\boldsymbol{\phi}$, γ^2, $\boldsymbol{\sigma}^2$, and $\mathbf{y}$.
2. Sample income process weights $\boldsymbol{\phi}$ conditional on $\boldsymbol{\omega}$, $\boldsymbol{\varepsilon}$, and $\mathbf{y}$.
3. Sample residual variance γ^2 conditional on $\boldsymbol{\phi}$, $\boldsymbol{\omega}$, $\boldsymbol{\varepsilon}$, and $\mathbf{y}$.
4. Sample volatility parameters $\sigma^2_{i,t}$ for each individual and year conditional on shocks $(\boldsymbol{\omega}, \boldsymbol{\varepsilon})$ and weights $\boldsymbol{\alpha}$.
5. Sample weight parameters $\boldsymbol{\alpha}$ conditional on volatility parameters $\sigma^2_{i,t}$.

In step 1, we sample the income process shocks $(\boldsymbol{\omega}, \boldsymbol{\varepsilon})$ while treating the income process parameters $\boldsymbol{\phi}$ and the entire distribution of income volatilities $\boldsymbol{\sigma}^2 = \{\sigma^2_{i,t}\}$ as fixed and known. In our state-space formulation of the income process (3), this step is equivalent to sampling the latent shocks $\mathbf{X}$ given fixed values of the parameter vector $\boldsymbol{\beta} = (1, \boldsymbol{\phi})'$ and the variance matrix $\boldsymbol{\Sigma}$, which contains volatility parameters $\sigma^2_{i,t}$. With known values of the parameters γ^2, $\boldsymbol{\beta}$, and $\boldsymbol{\Sigma}$, the Kalman filter (Kalman 1960) can be used to calculate maximum likelihood estimates of the latent shocks $\mathbf{X}$. Instead of focussing on point estimates, we use the Kalman filter within a Gibbs sampling algorithm outlined by Carter and Kohn (1994) to sample the full posterior distribution of the latent shocks $(\boldsymbol{\omega}, \boldsymbol{\varepsilon})$.

Step 2 of the Gibbs sampler, sampling our income process weights $\boldsymbol{\phi}$, becomes easy if we condition on our sampled latent shocks $(\boldsymbol{\omega}, \boldsymbol{\varepsilon})$ and residual variance γ^2, because the $\boldsymbol{\beta}$ vector containing our income process parameters acts in (3) as the coefficient vector for a regression model with known covariates and residual variance. Let $\mathbf{X}$ be the matrix that collects all vectors $\mathbf{X}_{i,t}$ across i and t, let $\mathbf{Y}$ be the vector that collects all excess log incomes $y_{i,t}$ across i and t, and let $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$ be the maximum likelihood estimate of $\boldsymbol{\beta}$. With flat priors on $\boldsymbol{\beta}$, the conditional posterior distribution of $\boldsymbol{\beta}$ is

$$\boldsymbol{\beta} \sim \text{MVNormal}[\hat{\boldsymbol{\beta}}, \gamma^2 \cdot (\mathbf{X}'\mathbf{X})^{-1}]$$

with the additional restriction that the $\boldsymbol{\phi}_{\varepsilon}$ elements of $\boldsymbol{\beta}$ must sum to 1.

Step 3 of this Gibbs sampler conditions on the sampled values of the income process shocks $(\boldsymbol{\omega}, \boldsymbol{\varepsilon})$ and the income process weights $\boldsymbol{\phi}$ to sample the residual variance γ^2. The conditional posterior distribution of γ^2 is

$$\gamma^2 \sim \text{Inv-Gamma}\left(\frac{N}{2}, \frac{(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})'(\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}})}{2}\right),$$

where N is the total number of income observations in our dataset, across all years and all people.

Step 4 of this Gibbs sampler generates new values for the volatility parameters $\boldsymbol{\sigma}^2 = (\sigma^2_{i,t})$ conditional on all of the other parameters and the observed data. Recall that each $\sigma^2_{i,t} \equiv (\sigma^2_{\varepsilon,i,t}, \sigma^2_{\omega,i,t})$ consists of both permanent and transitory volatility parameters for time t within individual i, so that these two values are sampled as a pair. To sample a full set of volatility parameters $\boldsymbol{\sigma}^2$ from $p(\boldsymbol{\sigma}^2|\boldsymbol{\phi}, \gamma^2, \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\varepsilon}, \mathbf{y})$, it is easiest to proceed sequentially by sampling (one-by-one) the volatility parameters $\sigma^2_{i,t}$ for individual i and year t from the distribution $p(\sigma^2_{i,t}|\boldsymbol{\phi}, \gamma^2, \boldsymbol{\alpha}, \boldsymbol{\omega}, \boldsymbol{\varepsilon}, \sigma^2_{-(i,t)}, \mathbf{y})$. $\sigma^2_{-(i,t)}$ represents the volatility parameters from other years within the individual as well as other individuals, so that we sample each volatility value conditioning on all other volatility values.

Noting that $\sigma^2_{i,t}$ is independent of $\boldsymbol{\phi}, \gamma^2$ and $\mathbf{y}$ given $\sigma^2_{-(i,t)}$ and the latent shocks $(\omega_{i,t}, \varepsilon_{i,t})$, the conditional posterior distribution $\sigma^2_{i,t}$ is

$$p\big(\sigma^2_{i,t}|\omega_{i,t}, \varepsilon_{i,t}, \sigma^2_{-(i,t)}, \boldsymbol{\alpha}\big) \propto p(\omega_{i,t}, \varepsilon_{i,t}|\sigma^2_{i,t}) \cdot p\big(\sigma^2_{i,t}|\sigma^2_{-(i,t)}, \boldsymbol{\alpha}\big). \quad (11)$$

The first term in (11) is the likelihood of our sampled shocks $(\omega_{i,t}, \varepsilon_{i,t})$ from our state-space model,

$$p(\omega_{i,t}, \varepsilon_{i,t}|\sigma^2_{i,t}) \propto (\sigma^2_{\omega,i,t} \cdot \sigma^2_{\varepsilon,i,t})^{-1/2} \exp\left(-\frac{1}{2}\frac{\omega^2_{i,t}}{\sigma^2_{\omega,i,t}} - \frac{1}{2}\frac{\varepsilon^2_{i,t}}{\sigma^2_{\varepsilon,i,t}}\right). \quad (12)$$

The second term in (11) comes from our MHDP prior described in Sections 2.3–2.4. We begin by sampling a volatility parameter proposal value ($\sigma^2_\star \equiv \{\sigma^2_{\omega,\star}, \sigma^2_{\varepsilon,\star}\}$) from the probability measure H given in (8). We set $\sigma^2_{i,t} = \sigma^2_\star$ only if we cannot find a suitable candidate $\sigma^2_{i,t} \in \sigma^2_{-(i,t)}$ among our currently existing values in the population.

Under our MHDP prior, the first candidate value for $\sigma^2_{i,t}$ that we consider is the immediately preceding value, $\sigma^2_{i,t-1}$. Recall our indicator variables $Z_{i,t} = 1$ if $\sigma^2_{i,t} = \sigma^2_{i,t-1}$ or $Z_{i,t} = 0$ otherwise. We need to sample a new value for $Z_{i,t}$ conditional on the shocks $\omega_{i,t}, \varepsilon_{i,t}$ and all other indicator variables $\mathbf{Z}_{i,-t}$ within individual i ($Z_{i,t}$ is independent of indicator variables $\mathbf{Z}_{j,t}$ for $j \neq i$),

$$\begin{aligned} p(Z_{i,t}|\omega_{i,t}, \varepsilon_{i,t}, \mathbf{Z}_{i,-t}, \boldsymbol{\alpha}) &\propto p(\omega_{i,t}, \varepsilon_{i,t}|Z_{i,t}) \cdot p(Z_{i,t}|\mathbf{Z}_{i,-t}, \boldsymbol{\alpha}) \\ &\propto p(\omega_{i,t}, \varepsilon_{i,t}|Z_{i,t}) \cdot p(Z_{i,t}|\mathbf{Z}_{i,1:t-1}, \boldsymbol{\alpha}) \\ &\quad \cdot \prod_{k=t+1}^{T} p(Z_{i,k}|Z_{i,t}, \ldots, Z_{i,k-1}, \boldsymbol{\alpha}). \end{aligned}$$

Thus the two values for $Z_{i,t}$ have relative probabilities

$$\begin{aligned} p(Z_{i,t} = 1) &\propto p(\omega_{i,t}, \varepsilon_{i,t}|\sigma^2_{i,t-1}) \cdot N^\star_{it} \\ &\quad \cdot \prod_{k=t+1}^{T} p(Z_{i,k}|Z_{i,t} = 1, \ldots, Z_{i,k-1}), \\ p(Z_{i,t} = 0) &\propto p(\omega_{i,t}, \varepsilon_{i,t}|\sigma^2_\star) \cdot \alpha_M \\ &\quad \cdot \prod_{k=t+1}^{T} p(Z_{i,k}|Z_{i,t} = 0, \ldots, Z_{i,k-1}). \end{aligned} \quad (13)$$

The first term in (13) is the likelihood of $(\omega_{i,t}, \varepsilon_{i,t})$ specified in equation (12). The second term in (13) comes from our Markovian prior specified in equation (9). The third term in (13) is calculated by carrying forward the prior specified in equation (9) for the subsequent years of individual i, conditional on either $Z_{i,t} = 1$ or $Z_{i,t} = 0$.

We sample $Z_{i,t}$ according to the probabilities in (13). Note that for $t = 1$ within each individual i, we skip the foregoing sampling step and set $Z_{i,1} = 0$. If we sample $Z_{i,t} = 1$, then we set $\sigma^2_{i,t} = \sigma^2_{i,t-1}$, and we are done with sampling $\sigma^2_{i,t}$.

If $Z_{i,t} = 0$, we then consider other values $\sigma^2_{i,-t}$ within individual i. Let k index the K_i unique volatility parameter values contained in $\sigma^2_{i,-t}$. We have $K_i + 1$ choices for $\sigma^2_{i,t}$ with probabilities

$$\begin{aligned} p(\sigma^2_{i,t} = \sigma^2_k) &\propto p(\omega_{i,t}, \varepsilon_{i,t}|\sigma^2_k) \cdot N_{ik}, \qquad k = 1, \ldots, K_i, \\ p(\sigma^2_{i,t} \notin \sigma^2_{i,-t}) &\propto p(\omega_{i,t}, \varepsilon_{i,t}|\sigma^2_\star) \cdot \alpha_H, \end{aligned} \quad (14)$$

where N_{ik} is the number of occurrences of the volatility parameter σ^2_k within the set of unique within-individual values $\sigma^2_{i,-t}$ and α_H is the prior weight for the group level of the HDP prior (7). The first term in (14) is again the likelihood of $(\omega_{i,t}, \varepsilon_{i,t})$ specified in equation (12). Note that the prior probability of selecting an existing within-individual value is proportional to its popularity (number of occurrences N_{ik}) within the individual. We sample one of these $K_i + 1$ choices for $\sigma^2_{i,-t}$ with relative probabilities given in (14). If we sample $\sigma^2_{i,t} = \sigma^2_k$, then we are done with sampling $\sigma^2_{i,t}$.

If we instead sample $\sigma^2_{i,t} \notin \sigma^2_{i,-t}$, then we next consider values for $\sigma^2_{i,t}$ from the current population of volatility parameters outside of the individual, which we denote as σ^2_{-i}. We now let σ^2_k denote the unique values in σ^2_{-i}, and let K denote the number of these unique values. We have $K + 1$ choices for $\sigma^2_{i,t}$ with probabilities

$$\begin{aligned} p(\sigma^2_{i,t} = \sigma^2_l) &\propto p(\omega_{i,t}, \varepsilon_{i,t}|\sigma^2_l) \cdot N_k, \qquad k = 1, \ldots, K, \\ p(\sigma^2_{i,t} \notin \sigma^2_{-i}) &\propto p(\omega_{i,t}, \varepsilon_{i,t}|\sigma^2_\star) \cdot \alpha_D, \end{aligned} \quad (15)$$

where N_k is the number of occurrences of σ^2_k within the set of volatility values σ^2_{-i} outside the individual. α_D is the prior weight for the population level of the HDP prior (8). Note that the prior probability of selecting an existing value is proportional to its popularity (i.e., number of occurrences N_l) across the population outside of the individual. We sample one of these $K + 1$ choices for $\sigma^2_{i,-t}$ with relative probabilities given in (15). If we sample $\sigma^2_{i,t} = \sigma^2_k$, then we are done with sampling $\sigma^2_{i,t}$.

If we instead sample $\sigma^2_{i,t} \notin \sigma^2_{-i}$, then we set $\sigma^2_{i,t} = \sigma^2_\star$, which is a new volatility value added to the population from measure H. This entire sequential procedure for sampling $\sigma^2_{i,t}$ for person i and year t then must be repeated for all other years and individuals to update our full set of volatility values $\boldsymbol{\sigma}^2$.

Finally, step 5 of our Gibbs sampler updates the weight parameters $\boldsymbol{\alpha} = (\alpha_M, \alpha_H, \alpha_D)$ conditional on current volatility parameters $\boldsymbol{\sigma}^2$. Each of these parameters are sampled using a Metropolis step. We sample a proposal value $\alpha^\star_M \sim \mathrm{N}(\alpha_M, c)$. The variance of the proposal distribution c is a tuning parameter ($c = 2$ worked well in our algorithm). This proposal value $\alpha^\star_M$

is accepted with probability $q = \min\{1, r\}$ where

$$r = \frac{p(\sigma^2|\alpha_M^\star) \cdot p(\alpha_M^\star)}{p(\sigma^2|\alpha_M) \cdot p(\alpha_M)}. \quad (16)$$

Otherwise, the current value α_M is retained. We update parameters α_H and α_D using an identical step.

Convergence of the Gibbs sampler for our full model (with Markovian extension) was diagnosed by running multiple chains for 20,000 iterations from well-dispersed starting points, as recommended by Gelman and Rubin (1992). In the current application, we observed that convergence had occurred after approximately 5000 iterations. The first 5000 iterations of several chains were discarded, and the remaining values were thinned (taking only every 20th iteration) to reduce autocorrelation.

3. APPLICATION TO LABOR INCOME DATA

Our data are the core sample of the PSID. The PSID was designed as a nationally representative panel of U.S. households (Hill 1991); it provides annual or biennial labor income spanning the years 1968 to 2005. Restricting ourselves to male household heads aged 22–60[7] gives us 52,181 observations on 3041 individuals with 17 years of recorded data per individual on average. After some initial processing and basic summaries presented in Section 3.1, we proceed to examine our main parameters of interest, the income process parameters and permanent and transitory income volatilities, in Section 3.4.

3.1 Initial Processing and Summaries

We focus on *excess* log income as our outcome measure, which is the residual from a regression to predict the natural log of labor income. This regression is weighted by PSID-provided sample weights, normalized so that the average weight in each year is the same. We use the following measures as covariates in this regression: a cubic in age for each level of educational attainment (none, elementary, junior high, some high school, high school, some college, college, graduate school); the presence and number of infants, young children, and older children in the household; the total number of family members in the household; and dummy variables for each calendar year. Including calendar year dummy variables eliminates the need to convert nominal income to real income explicitly.[8]

We want to ensure that changes in income are not driven by changes in the top code (i.e., the maximum value for income entered that can be entered in the PSID). The lowest top code for income was \$99,999 in 1982 (\$202,281 in 2005 dollars), after which the top code rises to \$9,999,999. To ensure that top codes are standardized in real terms, this minimum top code[9] is imposed on all years in real terms, so the top code is \$99,999 in 1982 and \$202,281 in 2005.

[7] Age restrictions are standard in the income dynamics literature, although exact age ranges vary slightly: Gottschalk and Moffitt (2002) (22–59), Meghir and Pistaferri (2004) (25–55), and Abowd and Card (1989) (21–64).

[8] Working with excess log real income is also standard in this literature (Carroll and Samwick 1997; Meghir and Pistaferri 2004).

[9] Our key results on volatility dynamics and heterogeneity are robust to other choices of this top code. Naturally, because changing the top code changes the range over which income can change, changing the top code does shift the distributions of income volatility slightly.

Table 1. Distribution of excess log income and income changes for men

	Excess income	
	Level	One-year change
Mean	0	0.0017
St. dev.	0.7307	0.4870
Observations	52,181	43,261
Minimum	−2.9325	−3.6877
5th percentile	−1.6283	−0.7323
25th percentile	−0.2964	−0.1089
50th percentile	0.1246	0.0134
75th percentile	0.4601	0.1442
95th percentile	0.9757	0.6673
Maximum	2.6435	3.5862

Because our income process does not model unemployment explicitly, we need to ensure that results for the log of income are not dominated by small changes in the level of income near 0 (which will imply huge or infinite changes in the log of income). To address this concern, we replace income values that are very small or 0 with a nontrivial lower bound. We choose as this lower bound the income that would be earned from a half-time job (1000 hours per year) at the real equivalent of the 2005 federal minimum wage (\$5.15 per hour). This imposes bottom codes of \$5150 in 2005 and \$2546 in 1982. Note that the difference in log income between the top and bottom codes is constant over time, so that differences in the prevalence of predictably extreme income changes over time cannot be driven by changes in the possible range of income changes. The vast majority of the values below this bound are exactly 0. This bound allows us to exploit transitions into and out of the labor force. Results are robust to other values for this lower bound, such as the income from full-time work (2000 hours per year) at the 2005 minimum wage (in real terms).

Table 1 presents the distribution of excess log income and 1-year changes in excess log income. The key thing to note is that most 1-year income changes are relatively modest, with half of all income changes (25th percentile through 75th percentile) between 11 percent and 14 percent (both in log points). However, a small fraction of income changes are extremely large, with the worst 5 percent of income changes below −73 percent and the best 5 percent better than 67 percent (again in log points). This provides suggestive evidence that a minority of extreme income changes represents much of the volatility in the data. To the degree that such extreme income changes are clustered within some individuals, it indicates substantial heterogeneity in income volatility, which we intend to model.

3.2 Addressing Missing Data

Some observations are missing within our dataset, mostly because no data were collected by the PSID in even-numbered years after 1997. We use an imputation step within our Markov chain Monte Carlo implementation that fills in these missing incomes with sampled values from their full conditional distribution under our model. Specifically, for a missing income value, $y_{i,t}$, we sample an imputed value $y_{i,t}^\star$ from the distribution of

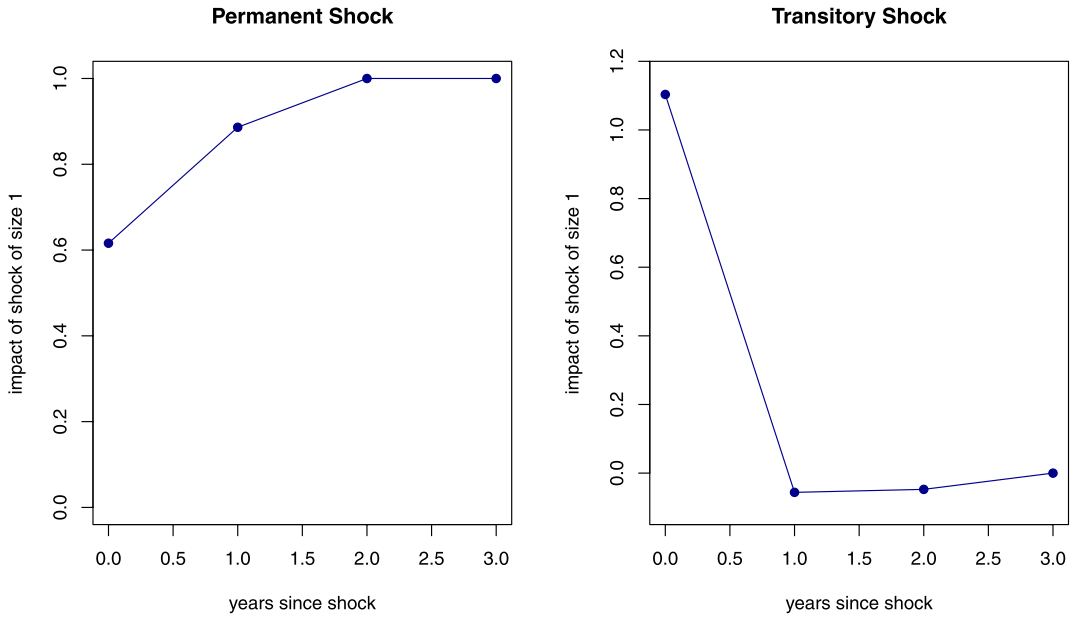

Figure 1. Impulse response function for permanent and transitory shocks. The online version of this figure is in color.

the missing value $y_{i,t}$ conditional on the surrounding observed values $y_{i,t-1}$ and $y_{i,t+1}$ and current values of all parameters.

The conditional distribution of $y_{i,t}$ given $y_{i,t+1} - y_{i,t-1} = z$ is normal, with mean and variance that are standard functions of z and current values of our volatility parameters $\boldsymbol{\sigma}^2$ and income process parameters $\boldsymbol{\phi}$. The imputation of all missing income values is repeated at the end of each full iteration of our Gibbs sampling algorithm. We also implemented our full model on a smaller dataset on data before 1997 in which all individuals containing missing values were removed and achieved similar results.

3.3 Estimated Parameters of Income Process

We first examine our posterior results for the global income process parameters of our model. Figure 1 shows the posterior means of the coefficients $\boldsymbol{\phi}$, which govern the rate at which shocks enter into and exit from income. These posterior means are presented in the form of an impulse response function, the impact of a 1 SD shock for an individual with average volatility parameters. As outlined in Section 2, we constrain $\boldsymbol{\phi}$ to be constant over time and across individuals. We see that the permanent shocks (left panel) enter in quickly, as indicated by $\phi_{\omega,k}$ increasing quickly to 1. We also see that transitory shocks (right panel) damp out quickly, as indicated by $\phi_{\varepsilon,k}$ quickly dropping to 0. Our model assumes that shocks enter and exit from income over three periods, but we estimate similar impulse response functions in models that allowed shocks to enter over two, four, or five periods. We estimate a 95% posterior interval of $(0.0037, 0.0040)$ for the global residual variance γ^2.

3.4 Estimated Volatility Parameters

The parameters of primary interest in our model are the parameters that govern volatility within individuals and over time. Figure 2 shows the distribution of the posterior mean of all permanent ($\boldsymbol{\sigma}^2_{\omega}$) and transitory ($\boldsymbol{\sigma}^2_{\varepsilon}$) volatility parameters, across all individuals and all years. Note the extreme skew and fat tails in the distribution of volatility parameters, $\boldsymbol{\sigma}^2$. We present both the full distributions (right plots) and the distributions with the right tail truncated (left plots), with that tail represented by a large mass at the extreme right of each plot.

Although the medians are modest, the means far exceed these medians. At the median, permanent shocks have an SD of approximately 16% annually; permanent shocks have an SD of approximately 18% annually. However, the upper tails of the transitory volatility parameters imply shocks with SDs well above 100% annually.

Figure 3 presents two features of the within-individual volatility distribution: the distribution of the number of volatility parameter values or clusters held by an individual [Figure 3(a)] and the probability that an individual's volatility value will be the same in 1, 2, 3, 4, or 5 years [Figure 3(b)]. At the median, an individual's volatility parameters take on seven values (changing six times over on average in 17 years of data). The probability that volatility remains unchanged after 5 years is roughly 20 percent. This indicates both that volatility is strongly persistent but also that the common assumption of constant volatility for an individual over time (as in Carroll and Samwick 1997) is violated.

We also examined the marginal posterior distributions of our weight parameters $\boldsymbol{\alpha}$. We estimate a 95% posterior interval of $(2, 5)$ for α_D, which weights the prior probability that volatility will change from the previous year within an individual in equation (13). We estimate a 95% posterior interval of $(17, 22)$ for α_H, which weights the prior probability of choosing unique volatility value within an individual in equation (14). Finally, we estimate a 95% posterior interval of $(31, 47)$ for α_D, which weights the prior probability of choosing unique volatility value across the population in equation (15).

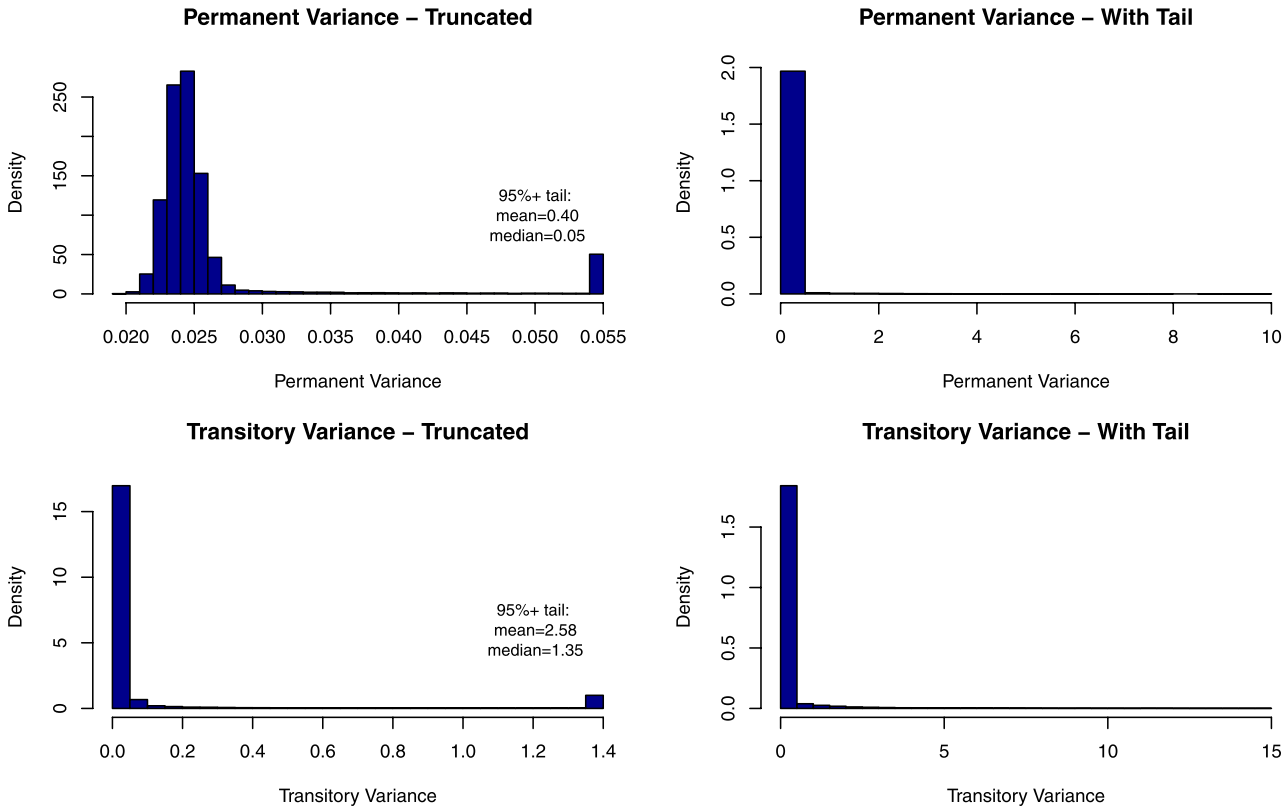

Figure 2. Distribution of posterior means of permanent variance $\sigma^2_{\omega,i,t}$ (top plots) and transitory variance $\sigma^2_{\varepsilon,i,t}$ (bottom plots) for each individual in each year. *Left*: Values are truncated at the 95th percentile for $\sigma^2_{\omega,i,t}$ and $\sigma^2_{\varepsilon,i,t}$. *Right*: All values are shown except for a small number of very extreme values (15 values of permanent variance fell between 10 and 19, whereas 28 values of transitory variance fell between 15 and 35). The online version of this figure is in color.

4. MODEL SENSITIVITY AND VALIDATION

4.1 Identifying Heterogeneity With Sample Moments

Our state-space model presented in Section 2.1 assumes that income shocks are normally distributed conditional on the volatility parameters (2). One consequence of this assumption is that unconditional kurtosis in the distribution of shocks is automatically attributed to heterogeneity in volatility parameters. However, an alternative hypothesis is that there is little or no heterogeneity in volatility parameters, but that the income shocks come from a more fat-tailed distribution than the normal distribution that we assume. When looking only at the cross-section of income changes, these two possibilities are observationally equivalent: heterogeneity in volatility parameters (with conditionally normal shocks) versus conditionally fat-tailed shocks (without heterogeneity in volatility parameters). However, by examining the serial dependence over time, it is possible to compare the two hypotheses. If shocks are conditionally fat-tailed but everyone has the same volatility parameters, then individuals with large past income changes should be no more likely than others to experience large subsequent income changes. If individuals differ in their volatility parameters and those volatilities are persistent, then individuals with large past income changes will be more likely than others to have large subsequent income changes.

We investigate these alternatives using the simplest sample moment that captures volatility, squared (2-year excess log) income changes:

$$(y_{it} - y_{it-2})^2.$$

In each year, we partition the data into two groups, individuals with and without large past absolute income changes, and compare the absolute size of subsequent income changes. In each year, a cohort without large past income changes is formed as the set of individuals whose squared income change was below the median 4 years ago. Correspondingly, a cohort with large past income changes is formed as the set of individuals whose squared change was above the 95th percentile 4 years ago. This 4-year period is chosen so that income shocks are far enough apart to be uncorrelated (Abowd and Card 1989). In Figure 4, we compare the distribution of squared income changes for the two cohorts. Note that current squared income changes are systematically and substantially lower for the cohort without large last squared income changes (solid line) than for those with large past squared income changes (dashed line). This indicates that volatility parameters are persistent, and that heterogeneity in estimated volatility parameters cannot be explained solely by fat-tailed but homogeneous shocks to income.

4.2 Stability of Income Process Parameters

In this section we compare estimates of income process weights $\boldsymbol{\phi}$ (governing the rate at which shocks enter into and

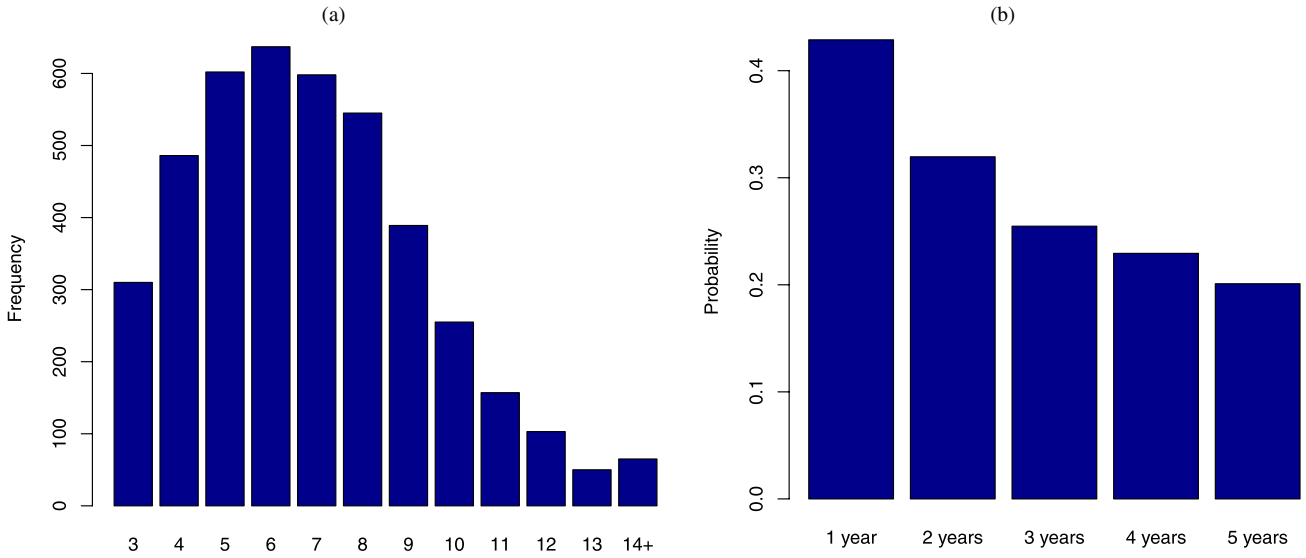

Figure 3. Distribution of volatility parameters within individuals. (a) Distribution of average number of volatility parameter values within individuals. (b) Empirical probability that an individual's volatility will be the same in 1, 2, 3, 4, or 5 years. The online version of this figure is in color.

exit from income) from our model to the same weights estimated under two simpler models of the volatility distribution. Specifically, in addition to our full MHDP model, we consider two alternatives: HDP, our model without the Markovian extension outlined in Section 2.4, and DP, an even simpler version of our model without the HDP outlined in Section 2.3. In each of these alternative models, we have different structures for our volatility parameters $\boldsymbol{\sigma}^2$ but the same state-space income process (3) parameterized by $\boldsymbol{\phi}$. Figure 5 shows boxplots of the posterior distribution of the income process weights $\boldsymbol{\phi}$ for our full model versus these simpler models. The distributions of $\boldsymbol{\phi}$ from each model show a reasonably similar pattern, suggesting that the income process parameters are not highly sensitive to different model choices for our volatility parameters $\boldsymbol{\sigma}^2$.

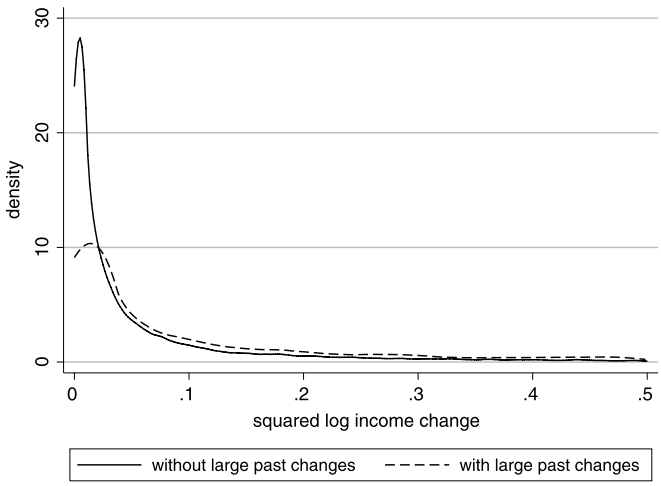

Figure 4. Distribution of squared income changes for those with and without large past income changes.

5. DISCUSSION

We have presented a sophisticated Bayesian hierarchical model for estimation of income volatility from panel data. Past approaches have focussed on overall summaries of income volatility across the population without addressing the heterogeneity between individuals in that population. Our semiparametric methodology based on an MHDP allows income volatility to vary between individuals and within individuals across time, while still sharing information over time and across individuals. We evaluated the validity of several model choices in Section 4, using sample moments to justify our modeling assumption of persistence in our distribution of volatility parameters $\boldsymbol{\sigma}^2$. Our income process parameters $\boldsymbol{\phi}$ did not seem to be sensitive to different models (DP vs. HDP vs. MHDP) for the income volatility parameters. Our methodological developments could be easily extended to other applications with grouped and ordered data, such as topic models where ordered e-mails are clustered based on word composition within a individual and across individuals (Zhu, Ghahramani, and Lafferty 2005).

Our semiparametric methodology leads to several interesting results when applied to data from the PSID. We estimate that the vast majority of observations are associated with modest volatility parameters (e.g., implying transitory income shocks with a standard deviation of 15% per year). However, a small minority have enormous income shock parameters (e.g., implying transitory income shocks with a standard deviation of $>100\%$ per year). We find that the distribution of volatility parameters is highly (positively) skewed with substantial excess kurtosis that would not be well approximated by a lognormal or χ^2 distribution, underscoring the importance of our flexible approach, which accommodates any shape of the volatility distribution.

[Received May 2009. Revised July 2011.]

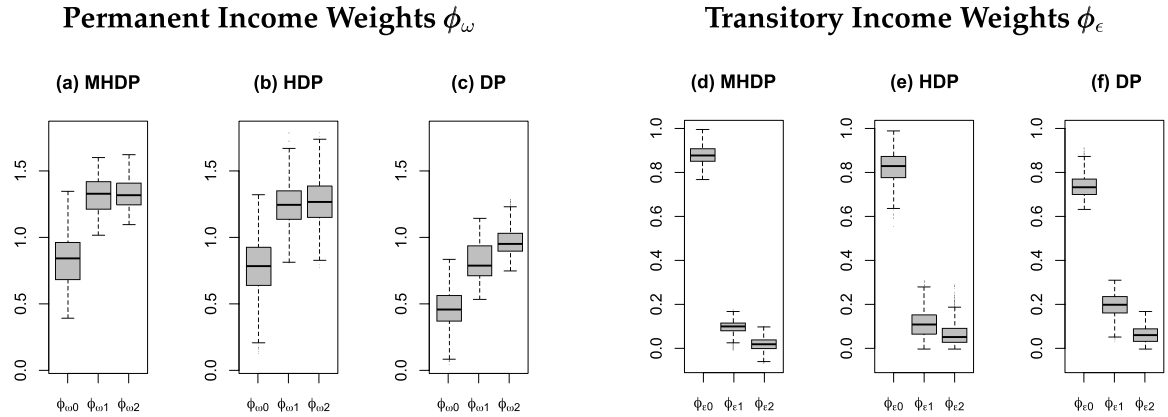

Figure 5. Distribution of the income process weights from the three models discussed in the text. Permanent income weights $\boldsymbol{\phi}_\omega$ are plotted in (a)–(c), and transitory income weights $\boldsymbol{\phi}_\epsilon$ are plotted in (d)–(f).

REFERENCES

Abowd, J., and Card, D. (1989), "On the Covariance Structure of Earnings and Hours Changes," *Econometrica*, 57, 411–445. [1281,1282,1286,1288]

Ahmed, A., and Xing, E. P. (2008), "Dynamic Non-Parametric Mixture Models and the Recurrent Chinese Restaurant Process," in *Proceedings of the Eighth SIAM International Conference on Data Mining (SDM2008)*, Atlanta, GA. [1283]

Baker, M. (1997), "Growth Rate Heterogeneity and the Covariance Structure of Life Cycle Earnings," *Journal of Labor Economics*, 15, 338–375. [1282]

Baker, M., and Solon, G. (2003), "Earnings Dynamics and Inequality Among Canadian Men, 1976–1992: Evidence From Longitudinal Income Tax Records," *Journal of Labor Economics*, 21, 289–321. [1282]

Banks, J., Blundell, R., and Grugiavini, A. (2001), "Risk Pooling and Consumption Growth," *Review of Economic Studies*, 68, 757–779. [1280-1282]

Blei, D. M., and Frazier, P. (2010), "Distance Dependent Chinese Restaurant Processes," in *Proceedings of the 27th International Conference on Machine Learning*, Haifa, Israel. [1284]

Blundell, R., Pistaferri, L., and Preston, I. (2008), "Consumption Inequality and Partial Insurance," *American Economic Review*, 98, 1887–1921. [1282]

Bollerslev, T. (1986), "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, 31, 307–327. [1281,1284]

Browning, M., Alvarez, J., and Ejrnaes, M. (2010), "Modelling Income Processes With Lots of Heterogeneity," *Review of Economic Studies*, 77 (4), 1353–1381. [1280,1282]

Carroll, C., and Samwick, A. (1997), "The Nature of Precautionary Wealth," *Journal of Monetary Economics*, 40, 41–71. [1281,1282,1286,1287]

Carter, C., and Kohn, R. (1994), "On Gibbs Sampling for State Space Models," *Biometrika*, 81, 541–553. [1280,1284]

Durbin, J., and Koopman, S. J. (2001), *Time Series Analysis by State Space Methods*, New York: Oxford University Press. [1282]

Engle, R. F. (1982), "Autoregressive Conditional Heteroscedasticity With Estimates of Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1008. [1281,1284]

Escobar, M. D., and West, M. (1995), "Bayesian Density Estimation and Inference Using Mixtures," *Journal of the American Statistical Association*, 90, 577–588. [1284]

Ferguson, T. (1974), "Prior Distributions on Spaces of Probability Measures," *The Annals of Statistics*, 2, 615–629. [1282,1283]

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2007), "The Sticky HDP-HMM: Bayesian Nonparametric Hidden Markov Models With Persistant States," Technical Report P-2777, MIT Laboratory for Information and Decision Systems. [1283]

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences," *Statistical Science*, 7, 457–472. [1286]

Geman, S., and Geman, D. (1984), "Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 6, 721–741. [1284]

Gottschalk, P., and Moffitt, R. (2002), "Trends in the Transitory Variance of Earnings in the United States," *Economic Journal*, 112, C68–C73. [1282, 1286]

Griffin, J., and Steel, M. (2006), "Order-Based Dependent Dirichlet Processes," *Journal of the American Statistical Association*, 101, 179–194. [1283]

Hill, M. S. (1991), *The Panel Study of Income Dynamics: A User's Guide*, Vol. 2, Thousand Oaks, CA: Sage. [1286]

Hirano, K. (2002), "Semiparametric Bayesian Inference in Autoregressive Panel Data Models," *Econometrica*, 70, 781–799. [1283]

Iorio, M. D., Muller, P., Rosner, G. L., and MacEachern, S. N. (2004), "An ANOVA Model for Dependent Random Measures," *Journal of the American Statistical Association*, 99, 205–215. [1283]

Jensen, S. T., and Liu, J. S. (2008), "Bayesian Clustering of Transcription Factor Binding Motifs," *Journal of the American Statistical Association*, 103, 188–200. [1280]

Jensen, S. T., and Shore, S. H. (2009), "Changes in the Distribution of Income Volatility," technical report, The Wharton School, University of Pennsylvania, Philadelphia, PA. Available at *arXiv:0808.1090*. [1280]

Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," *Transaction of the ASME—Journal of Basic Engineering*, 82, 35–45. [1280,1284]

MacEachern, S. N. (1999), "Dependent Nonparametric Processes," in *Proceedings of the Section on Bayesian Statistical Science*, Alexandria, VA: American Statistical Association. [1283]

Medvedovic, M., and Sivaganesan, S. (2002), "Bayesian Infinite Mixture Models Based Clustering of Gene Expression Profiles," *Bioinformatics*, 18, 1194–1206. [1280]

Meghir, C., and Pistaferri, L. (2004), "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72, 1–32. [1280-1282,1284,1286]

Meghir, C., and Windemejier, F. (1999), "Moments Conditions for Dynamic Panel Data Models With Multiplicative Individual Effects in the Conditional Variance," *Annales d'Economie et de Statistique*, 55–56, 317–330. [1280-1282]

Muller, P., and Quintana, F. A. (2004), "Nonparametric Bayesian Data Analysis," *Statistical Science*, 19, 95–110. [1280]

Quintana, F. A., Mueller, P., Rosner, G. L., and Munsell, M. (2008), "Semi-Parametric Bayesian Inference for Multi-Season Baseball Data," *Bayesian Analysis*, 3, 317–338. [1280]

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2008), "The Nested Dirichlet Process," *Journal of the American Statistical Association*, 103, 1131–1154. [1283]

Shin, D., and Solon, G. (2008), "Trends in Men's Earnings Volatility: What Does the Panel Study of Income Dynamics Show?" Working Paper 14075, National Bureau of Economic Research. [1282]

Teh, Y., Jordan, M., Beal, M., and Blei, D. (2006), "Hierarchical Dirichlet Processes," *Journal of the American Statistical Association*, 101, 1566–1581. [1281,1283]

Wang, X., and McCallum, A. (2006), "Topics Over Time: A Non-Markov Continuous-Time Model of Topical Trends," in *KDD 06*, New York: Association for Computing Machinery. [1283]

Xue, Y., Dunson, D., and Carin, L. (2007), "The Matrix Stick-Breaking Process for Flexible Multi-Task Learning," in *Proceedings of the 24th International Conference on Machine Learning (ICML 2007)*, New York: Association for Computing Machinery. [1283]

Zhu, X., Ghahramani, Z., and Lafferty, J. (2005), "Time-Sensitive Dirichlet Process Mixture Models," technical report, School of Computer Science, Carnegie Mellon University. [1283,1284,1289]