Chapter 1

Predictive Distributions based on Longitudinal Earnings Data

Gary Chamberlain

Keisuke Hirano

1.1 Introduction

Consider an individual trying to forecast his future earnings, in order to guide savings and other decisions. How should he proceed? We shall focus on how he might combine his personal earnings history with data on the earnings trajectories of other individuals. A key issue is the sense in which the experiences of others are relevant for him.

Here is a stylized version of the problem. At the beginning of each period, the individual has some amount of financial wealth. He receives labor earnings and pays out-of-pocket medical expenses. Financial wealth plus labor earnings minus medical expenses gives cash on hand. He chooses to consume some of this, with the rest invested in various assets. Summing up the returns on these assets gives financial wealth at the beginning of the next period, and the process repeats. The individual faces various sorts of uncertainty. Labor income in future periods is uncertain due, for example, to unemployment risk. The real returns on at least some of the assets are uncertain, and there is uncertainty about future medical expenses. A decision rule specifies consumption at each date as a function of cash on hand at that date and of variables (whose values are known at that date) that are used in forming conditional distributions for the future outcomes. Such a decision rule leads to a distribution for the consumption stream, and the individual uses expected utility preferences to rank the distributions corresponding to different decision rules. The objective is to choose a decision rule that maximizes expected utility.

Recent work on this problem includes Skinner (1988), Caballero (1990), Deaton (1991), Hubbard, Skinner, and Zeldes (1994, 1995), and Carroll (1997). These papers adopt specifications for preferences and for the conditional distribution of the uncertain outcomes. They use analytical and numerical methods to solve for optimal decision rules, and then summarize properties of the optimal paths for consumption and for the stock of financial wealth.

Our interest is in the specification of the conditional distribution for earnings. We envision an individual seeking advice from a financial planner. The individual provides data on his earnings history and on various personal characteristics such as age and education. The planner has access to longitudinal data sets that provide data on earnings histories and personal characteristics for samples of individuals. We would like to devise optimal ways to combine the individual's information with the survey data in order to provide the individual with a conditional distribution of his future earnings.

1.2 Predictive Distributions

There are observations on the earnings of N individuals over T years: $\{y_i^T\}_{i=1}^N$, where $y_i^T := \{y_{it}\}_{t=1}^T$. In addition, data are available on some characteristics of these sample individuals, such as their age and education: $\{x_i\}_{i=1}^N$. This data is available to our decision maker. The observations on earnings are regarded as partial realizations of the random variables $\{Y_i^H\}_{i=1}^N$, where $Y_i^H := \{Y_{it}\}_{t=1}^H$ and $H \geq T$. The random variables $\{Y_i^H\}_{i=1}^N$ are distributed according to P_θ for some value of the parameter θ in the parameter space Θ. P_θ specifies that these random variables are independent, and that the distribution of Y_i^H has density $f_H(\cdot \,|\, x, \theta)$ with $x = x_i$. This implies a marginal density, $f_s(\cdot \,|\, x, \theta)$, for Y_i^s with $x = x_i$ and $s \leq H$.

Let d denote our decision maker. He wants to construct a distribution for his future earnings that could be used in expected utility calculations. He knows his earnings history y_d^K and characteristics x_d, and wants to construct a joint distribution for his future earnings $Y_{d,K+1}^H := \{Y_{d,t}\}_{t=K+1}^H$, conditional on $\{Y_d^K = y_d^K\}$. Suppose that he regards himself as conditionally exchangeable with the sample individuals, in that his earnings are realizations from the same distribution conditional on characteristics: $Y_d^H \sim f_H(\cdot \mid x_d, \theta)$. Then our decision maker d could have been part of the sample, and we shall also refer to him as $N+1$ (with the independence assumption extended to $\{Y_i^H\}_{i=1}^{N+1}$). Given θ, the distribution of his future earnings has density $f_{H|K}(\cdot \mid y_d^K, x_d, \theta)$, where

$$f_{H|K}(v \mid u, x, \theta) = f_H(u, v \mid x, \theta)/f_K(u \mid x, \theta).$$

This distribution is not immediately useful for decision making, however, because θ is unknown.

Our decision maker (or his agent, the financial planner) can deal with the unknown θ by introducing a prior distribution with density $p(\theta)$. The data consist of

$$z = (\{y_i^T, x_i\}_{i=1}^N, y_{N+1}^K, x_{N+1}).$$

There is a posterior distribution with density

$$p(\theta \mid z) \propto \prod_{i=1}^N f_T(y_i^T \mid x_i, \theta) f_K(y_{N+1}^K \mid x_{N+1}, \theta) p(\theta).$$

This can be used to calculate the following predictive density:

$$f(y_{K+1}, \ldots, y_H \mid y_d^K, x_d, z) = \int f_{H|K}(y_{K+1}, \ldots, y_H \mid y_d^K, x_d, \theta) p(\theta \mid z)\, d\theta. \tag{1}$$

We shall evaluate this predictive distribution numerically for particular histories y_d^K and characteristics x_d.

1.3 The Model

Our choice of model is largely based on the previous literature on longitudinal earnings data, including Hause (1977), Lillard and Willis (1978), Chamberlain (1978), MaCurdy (1982), Abowd and Card (1989), Card (1994), Moffitt and Gottschalk (1995), and Geweke and Keane (1997). Our model has one feature, however, that appears not to have been examined before; it allows for heterogeneity across individuals in the variance of the earnings innovations.

$$
\begin{aligned}
Y_{it} &= g_t(x_i, \beta) + V_{it} + \alpha_i + U_{it} \qquad (t = 1, \ldots, H) \qquad\qquad (M)\\
V_{it} &= \gamma V_{i,t-1} + W_{it} \qquad (t = 2, \ldots, H),
\end{aligned}
$$

where

$$
\begin{aligned}
&V_{i1} \sim \mathcal{N}(0, \sigma_v^2), \quad W_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2), \quad \alpha_i \,|\, V_{i1} \sim \mathcal{N}(\psi V_{i1}, \sigma_\epsilon^2)\\
&h_i \sim \mathcal{G}(m/2, \tau/2), \quad U_{it} \,|\, h_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, h_i^{-1}).
\end{aligned}
$$

($\mathcal{G}(r, s)$ denotes a gamma distribution with mean r/s and variance r/s^2.) Furthermore, (V_{i1}, α_i), $\{W_{it}\}_{t=2}^H$, and $(h_i, \{U_{it}\}_{t=1}^H)$ are mutually independent (under P_θ). The $g_t(\cdot, \cdot)$ are given functions, and the parameter vector is $\theta = (\beta, \gamma, \psi, \sigma_v, \sigma_w, \sigma_\epsilon, m, \tau)$. The individual is assumed to begin working in period $t = 1$.

An Inconsistent Bayes Estimator. We shall briefly consider the consequences of an alternative prior distribution, in which the $\{\alpha_i\}_{i=1}^N$ are treated as parameters (part of θ)

and assigned a uniform (improper) prior. This can be a very poor choice, and we shall use a simple model to make this point. We observe $Z := \{Z_i\}_{i=1}^N \sim P_\theta$ for some $\theta \in \Theta$. $Z_i = (Y_{i1}, \ldots, Y_{iT})$ and P_θ specifies that

$$Y_{it} = \gamma Y_{i,t-1} + \alpha_i + W_{it}$$

$$W_{it} \,|\, Y_1 \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \qquad (i = 1, \ldots, N;\, t = 2, \ldots, T),$$

where $Y_1 = \{Y_{i1}\}_{i=1}^N$. The conditional density of $(Y_{i2}, \ldots, Y_{iT})$ given $Y_{i1} = y_{i1}$ is

$$f(y_{i2}, \ldots, y_{iT} \,|\, y_{i1}, \theta) \,\propto\, \sigma^{-(T-1)} \exp[-\frac{1}{2\sigma^2} \sum_{t=2}^{T} (y_{it} - \gamma y_{i,t-1} - \alpha_i)^2].$$

The conditional density of Z given $\{Y_{i1} = y_{i1}\}_{i=1}^N$ is

$$f(z \,|\, y_1, \theta) \,\propto\, \sigma^{-n} \exp[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\lambda)'(\mathbf{y} - \mathbf{X}\lambda)],$$

where $n = N(T-1)$ and

$$y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Nt} \end{pmatrix}, \quad \mathbf{y} = \begin{pmatrix} y_2 \\ \vdots \\ y_T \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} y_1 & I_N \\ \vdots & \vdots \\ y_{T-1} & I_N \end{pmatrix}, \quad \lambda = \begin{pmatrix} \gamma \\ a_1 \\ \vdots \\ a_N \end{pmatrix}.$$

Let $f_1(y_1 \,|\, \theta)$ denote the P_θ density for Y_1. If the prior density for θ is $p(\theta)$, then the posterior density is

$$p(\theta \,|\, z) \,\propto\, \sigma^{-n} \exp[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\lambda)'(\mathbf{y} - \mathbf{X}\lambda)] f_1(y_1 \,|\, \theta) p(\theta).$$

Let $\theta = (\theta_1, \theta_2)$, where $\theta_1 = (\lambda, \sigma)$. Suppose that $\Theta = \Theta_1 \times \Theta_2$, $f_1(y_1 \,|\, \theta) = f_1(y_1 \,|\, \theta_2)$, and $p(\theta_1, \theta_2) = p_1(\theta_1) p_2(\theta_2)$. In this case, Y_1 is uninformative for θ_1. If $p_1(\lambda, \sigma) \propto \sigma^{-1}$, then the posterior density for θ_1 is

$$p_1(\theta_1 \,|\, z) \,\propto\, \sigma^{-n-1} \exp[-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}\lambda)'(\mathbf{y} - \mathbf{X}\lambda)].$$

This is identical to the posterior density that arises in the classical regression model ($Y \,|\, \mathbf{X} \sim \mathcal{N}(\mathbf{X}\lambda, \sigma^2 I)$) with a diffuse prior:

$$\lambda \,|\, \sigma, z \;\sim\; \mathcal{N}(\hat{\lambda}, \sigma^2(\mathbf{X}\mathbf{X})^{-1}), \quad \hat{\lambda} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\sigma^{-2} \,|\, z \;\sim\; \chi^2(n - N - 1)/s^2, \quad s^2 = (\mathbf{y} - \mathbf{X}\hat{\lambda})'(\mathbf{y} - \mathbf{X}\hat{\lambda}).$$

The standard residual regression argument shows that

$$\gamma \,|\, \sigma, z \;\sim\; \mathcal{N}(\hat{\gamma}, \sigma^2 W^{-1}),$$

where $\hat{\gamma}$ is the within estimator and W is the within sum of squares for $y_{i,t-1}$:

$$W = \sum_{i=1}^{N} \sum_{t=2}^{T} (y_{i,t-1} - \bar{y}_{i,-1})^2, \quad \bar{y}_{i,-1} = \frac{1}{T-1} \sum_{t=2}^{T} y_{i,t-1},$$

$$\hat{\gamma} = W^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} (y_{i,t-1} - \bar{y}_{i,-1})(y_{it} - \bar{y}_i), \quad \bar{y}_i = \frac{1}{T-1} \sum_{t=2}^{T} y_{it}.$$

Suppose that $T = 3$. Then the within estimator is obtained from a first-difference regression:

$$\hat{\gamma} = \sum_{i=1}^{N} (y_{i2} - y_{i1})(y_{i3} - y_{i2}) \Big/ \sum_{i=1}^{N} (y_{i2} - y_{i1})^2.$$

Since

$$Y_{i3} - Y_{i2} = \gamma(Y_{i2} - Y_{i1}) + W_{i3} - W_{i2},$$

the sampling distribution of $\hat{\gamma}$ gives

$$\mathrm{plim}_{N\to\infty}\, \hat{\gamma} = \gamma - \sigma^2 / [\mathrm{plim}\, \frac{1}{N} \sum_{i=1}^{N} (Y_{i2} - Y_{i1})^2].$$

So the Bayes estimator of γ under a symmetric loss function is inconsistent as $N \to \infty$.

If $Y_{i1} = 0$ for all i, then it is appropriate to regard Y_1 as uninformative for (λ, σ), as this analysis does. But note that $\text{plim}\, \frac{1}{N}\sum_{i=1}^{N}(Y_{i2} - Y_{i1})^2 = \lim \frac{1}{N}\sum_{i=1}^{N}\alpha_i^2 + \sigma^2$ in this case; the inconsistency of $\hat{\gamma}$ vanishes as $\frac{1}{N}\sum_{i=1}^{N}\alpha_i^2 \to \infty$, which is exactly what is implied by the improper, uniform prior on $\{\alpha_i\}_{i=1}^{N}$. The (finite-sample) optimality of the Bayes estimator is not being contradicted—the risk properties (in repeated samples) are fine if we average the risk function over Θ using the prior, which will focus on the part of the parameter space where $\frac{1}{N}\sum_{i=1}^{N}\alpha_i^2$ is very large.

One response is that this is not a relevant part of the parameter space, and we should consider prior distributions that do not dogmatically assume that $\frac{1}{N}\sum_{i=1}^{N}\alpha_i^2$ is very large. The uniform prior for $\{\alpha_i\}_{i=1}^{N}$ is in fact very informative when N is large, and a lot of data (large T) may be required to dominate it. An alternative prior could have a hierarchical (nested) form, such as $\alpha_i \,|\, \mu, \sigma_\epsilon \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma_\epsilon^2)$ with a quite diffuse prior distribution for μ, σ_ϵ. The uniform prior for α corresponds to setting $\sigma_\epsilon = \infty$; better to put a prior on σ_ϵ and let the data help to determine its value. This is the approach that we take in the empirical work.

1.4 Results

Our sample is drawn from the Panel Study of Income Dynamics (PSID). We work with a sample of males who were heads of household at ages 24 through 33. The earnings measure begins in the calendar year prior to the interview when the individual was 25 years old. We assume that "serious" labor force attachment begins in that year ($t = 1$). The selection criteria are that the sample individuals be in the labor force, not in school, with positive earnings for the $T = 10$ consecutive years beginning with $t = 1$. In addition, there should be no missing values for race or education. This gives $N = 813$ individuals. Note that $t = 1$ may refer to a different calendar year (between 1967 and 1991) for different individuals. We shall also work with a subset of this sample consisting of $N = 516$ whites who were not part of the Survey of Economic Opportunity (SEO) sample, which oversampled poor households. This group in turn is broken down into three disjoint subgroups, consisting of individuals with 9-11 years of schooling (high school dropouts), high school graduates, and college graduates.

Y_{it} is the logarithm of annual labor earnings for individual i in his t^{th} year in the sample. The earnings are in 1991 dollars, based on the urban consumer price index. The characteristics x_i consist of a constant, an indicator variable for race (1 if white, 0 otherwise) and five indicator variables based on highest grade completed: 9-11, high school graduate, high school plus some nonacademic training or some college (no degree), bachelor's degree, advanced or professional degree. In order to focus on the serial correlation structure of earnings residuals, we simply set $g_t(x_i, \beta)$ equal to $x_i'\beta_t$, and estimate the β_t's using separate least-squares regressions for each year t of experience. We then treat $\beta' = (\beta_1', \ldots, \beta_T')$ as known and equal to these estimates, and work with the residuals $Y_{it} - g_t(x_i, \beta)$. For the rest of the paper, we shall simply let Y_{it} denote these residuals.

The prior distribution for the parameters of the residual process is chosen to be relatively uninformative. The priors for γ, ψ, and $\log \sigma_w$ are (improper) uniform. The prior for m is uniform on a discrete grid extending from .1 to 25 with steps of size .01.

The priors for σ_v^{-2}, σ_ϵ^{-2}, and τ are $\chi^2(1)/.01$. (So the prior .025 and .975 quantiles for σ_v, for example, are .04 and 3.20.) All these components of θ are independent in the prior.

We shall use posterior medians as point estimates. These are reported for the full sample ($N = 813$) in Table 1.1, along with the .025 and .975 quantiles of the posterior distributions. The estimate for the autoregressive parameter γ is .98. This is sufficiently close to one to make it difficult to distinguish between the time-invariant component α_i and the initial condition V_{i1} for the autoregressive process $\{V_{it}\}_{t=1}^{T}$. This need not be a problem, however, since our main interest is in the predictive distribution. If different combinations of σ_ϵ and σ_v lead to similar predictive distributions for future earnings, then we do not need to distinguish between them.

We are particularly interested in the posterior distribution of (m, τ) and its implications for the predictive distribution. It may be easier to think about magnitudes if we work with $\sigma_{ui} := 1/\sqrt{h_i}$, the conditional standard deviation of the transitory shocks U_{it}. A value for (m, τ) implies a $\mathcal{G}(m/2, \tau/2)$ distribution for h, which in turn implies a distribution for σ_u. Let $\bar{\sigma}_u$ denote the median of this distribution; so $\bar{\sigma}_u$ is a function of (m, τ). The posterior distribution for (m, τ) implies a posterior distribution for $\bar{\sigma}_u$, and Table 1.1 reports the median, .025, and .975 quantiles of this posterior distribution. The estimate of .11 is comparable in size to the .10 estimate for the standard deviation σ_w of the innovations W_{it} in the autoregressive process.

We can obtain a distribution for h, and hence for σ_u, by taking the $\mathcal{G}(m/2, \tau/2)$ distribution, which conditions on (m, τ), and integrating with respect to the posterior distribution for (m, τ). This gives the predictive distribution for h; i.e., the posterior distribution for h_{N+1}, corresponding to an individual who was not in the sample. Quantiles of the resulting distribution for σ_u are in Table 1.2, and there is a plot of the density in Figure 1.1. The distribution is highly skewed and dispersed; the .05, .50, and .95 quantiles are .04, .11, and .81. This corresponds to the posterior distribution of m being concentrated on the interval $(1.1, 1.3)$.

The posterior distributions for the parameters θ can now be combined with an earnings history y_d^K to form a predictive distribution. We would like to describe these predictive distributions for an interesting range of earnings histories. We have selected eight individuals from the sample based on their means $(\bar{y}_i = \Sigma_{t=1}^{T} y_{it}/T)$, last observation (y_{iT}), and standard deviations $(\mathrm{std}(y_i) = (\Sigma_{t=1}^{T}(y_{it} - \bar{y}_i)^2/(T-1))^{1/2})$ for residual log(earnings). The results are in Table 1.3 and in Figures 1.2–1.5. First consider sample individuals $i = 38$ and $i = 346$. They are chosen to provide a contrast in the means, with similar values for the last observation and the standard deviation. The sample means are $-.66$ and $.51$, which fall near the .1 and .9 quantiles of the sample distribution of $\bar{y}_i$. Table 1.3 shows the quantiles of the predictive distributions for $Y_{i,T+h}$ at horizons $T+1$, $T+5$, $T+10$, and $T+20$. Figure 1.2 plots the sample data and the quantiles of the predictive distributions for the log(earnings) residuals. There are also plots of the predictive densities. The results indicate that the locations of the predictive distributions are very sensitive to the individual sample histories, and the locations show considerable persistence across the different horizons. The median of the predictive distribution for $i = 38$ is $-.58$ at $T+1$. The median shifts towards zero as the horizon increases but is still $-.42$ at $T+20$. Likewise for $i=346$, the median is .42 at $T+1$, shifts towards zero as the horizon increases, but is still .31 at $T+20$.

Our second contrast is between individuals $i=196$ and 491, who differ in their final observation: $-.93$ vs. .65. The locations of the predictive distributions are quite sensitive to this difference in the sample histories. The median for $i=196$ is $-.35$ at $T+1$; for $i=$ 491, it is .49. There is considerable persistence: at $T+20$ the medians are $-.24$ and .34.

Next consider two individuals ($i=321$ and 415) who differ in their sample standard deviations: .07 vs. .47 (which are near the .1 and .9 quantiles of the distribution of $\text{std}(y_i)$). The predictive distributions at $T+1$ show a sharp difference in spreads. The difference between the .90 and .10 quantiles is .32 for the low standard deviation person and 1.19 for the high standard deviation person. This contrast erodes over time, since the spread of the predictive distribution increases much more with horizon for the first individual. At the twenty year horizon, the 90-10 differentials are 1.02 and 1.52.

Our final contrast is between two individuals ($i=481$ and 297) who differ in their sample means (.12 vs. -1.07) and standard deviations (.07 vs. .56). The location of the predictive distribution for the second individual is sensitive to his low sample mean; the median of the predictive distribution is $-.91$ at $T+1$, and shifts towards zero as the horizon increases, but is still $-.66$ at $T+20$. The spreads of the predictive distributions are very different at $T+1$, with 90-10 differentials of .33 vs. 1.60. But the spread increases sharply with the horizon for $i=481$, whereas it is fairly constant for $i=297$—see Figure

1.5. So by $T+20$, the 90-10 differentials are 1.03 and 1.85.

These calculations have been repeated for a subsample of whites who were not in the SEO portion of the PSID. Table 1.1 reports the parameter estimates for three separate education groups, based on highest grade completed. There are 37 high-school dropouts (grades 9–11), 100 high-shool graduates, and 122 college graduates. The estimates of the autoregressive parameter γ for the three groups are similar to the estimate for the full sample: .95, .97, .99. The predictive distributions for σ_u in Table 1.2 and Figure 1.1 are also similar to the full sample results, showing a great deal of skewness and dispersion. The main differences are in the point estimates of σ_v and σ_ϵ, but their .95 intervals are wide and there is a great deal of negative dependence in the joint posterior distribution of $(\sigma_v, \sigma_\epsilon)$. In fact, the predictive distributions for Y based on the subsamples and the full sample are quite similar, a point we shall return to below.

1.4.1 *Normal Model with No Heterogeneity in Volatility*

If m and τ tend to infinity with m/τ fixed at $\bar{h}$, then the distribution of h becomes a point mass at $\bar{h}$, and our model reduces to one with no heterogeneity in the volatility. Our results do not support such a reduction, since the posterior distribution of m is concentrated between 1.07 and 1.33. A value for (m, τ) implies a distribution for $\sigma_u = 1/\sqrt{h}$; let *iqr* denote the interquartile range of this distribution. The (central) posterior .95 interval for *iqr* is (.12, .16), implying that the h distribution is not degenerate. Nevertheless, we have fit the model with h_i restricted to equal a constant, $\bar{h}$, in order to see how much difference this makes in the predictive distributions. The prior on $\bar{h}$ is $\chi^2(1)/.01$. The posterior medians and .95 probability intervals are as follows: γ: .73 (.66, .85); ψ: $-.35$ ($-.44$, $-.15$); σ_v: .34 (.30, .38); σ_w: .20 (.18, .21); σ_ϵ: .33 (.27, .35); σ_u: .18 (.16, .19). ($\sigma_u \equiv 1/\sqrt{\bar{h}}$.) Table 1.4 provides quantiles of the predictive distributions for four of the sample individuals considered above, and Figures 1.6 and 1.7 plot the quantiles and the predictive densities.

Individuals $i = 38$ and 346 differ in their sample means ($-.66$ vs. .51). The medians of the predictive distributions for $i = 38$ at $T+1$ and $T+20$ are $-.38$ and $-.45$. In Table 1.3, based on the more general model, we had $-.58$ and $-.42$. For $i = 346$, the $T+1$ and $T+20$ medians are .33 and .37 in Table 1.4; they are .42 and .31 in Table 1.3. So the restricted model has less sensitivity to $\bar{y}_i$ in the location of the predictive distribution at $T+1$. But this sensitivity increases with the horizon, instead of decreasing as in the more general model.

Now consider individuals $i = 481$ and 297, who differ in their sample means (.12 vs. -1.07) and standard deviations (.07 vs. .56). The medians of the predictive distribution for $i = 297$ at $T+1$ and $T+20$ are $-.72$ and $-.76$. In Table 1.3 we had $-.91$ and $-.66$. So again the restricted model shows less location sensitivity at $T+1$ than the more general model, but this reverses at the longer horizon.

There is a very sharp contrast across the two models in the spreads of the predictive distributions. In the more general model, the 90-10 differentials at $T+1$ for $i = 481$ and $i = 297$ are .33 and 1.60 (Table 1.3 and Figure 1.5). In the restricted model, these spreads are .75 and .75 (Table 1.4 and Figure 1.7). Under multivariate normality, the conditional distribution of $Y_{N+1,T+h}$, given θ and the earnings history $(y_{N+1,1},\ldots,y_{N+1,T})$, is normal with a conditional variance that does not depend upon the earnings history—only the location of the distribution is affected by the earnings history. This is reflected in the spread of an individual's predictive distribution not being sensitive to the amount of volatility in the individual's earnings history.

1.5 Extensions

1.5.1 *A Variance-Components Model for Volatility*

Our main finding is that there is substantial heterogeneity in volatility, and that this has important consequences for forming predictive distributions. We want to check whether this result holds up in more general versions of the model. In particular, we are concerned that model (M) uses heterogenetiy in volatility to capture other aspects of nonnormality. Consider the following extension:

$$\begin{aligned} Y_{it} &= V_{i,t} + \alpha_i + U_{it} \qquad (t=1,\ldots,H) \\ V_{it} &= \gamma V_{i,t-1} + W_{it} \qquad (t=2,\ldots,H), \end{aligned} \tag{M_1}$$

where

$$\begin{aligned} &V_{i1} \sim \mathcal{N}(0,\sigma_v^2), \quad W_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,\sigma_w^2), \quad \alpha_i \,|\, V_{i1} \sim \mathcal{N}(\psi V_{i1}, \sigma_\epsilon^2) \\ &h_{it} = \phi_i \zeta_{it}, \quad \zeta_i = \{\zeta_{it}\}_{t=1}^H, \quad U_{it} \,|\, \phi_i, \zeta_i \overset{\text{ind}}{\sim} \mathcal{N}(0, h_{it}^{-1}) \\ &\phi_i \sim \mathcal{G}(m/2, \tau/2), \quad \zeta_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(m_1/2, 1/2). \end{aligned}$$

(Recall that β in the term $g_t(x_i,\beta)$ in model (M) is being treated as known, and Y_{it} now denotes the residual.) Furthermore, (V_{i1},α_i), $\{W_{it}\}_{t=2}^H$, and $(\phi_i,\zeta_i,\{U_{it}\}_{t=1}^H)$ are mutually independent under P_θ; also, ϕ_i and ζ_i are independent. Now the parameter vector is $\theta = (\gamma, \psi, \sigma_v, \sigma_w, \sigma_\epsilon, m, \tau, m_1)$—there is an additional parameter m_1.

This model specializes to our original model (M) if the ζ_{it} distribution (appropriately scaled) is degenerate. This occurs if $m_1 \to \infty$ with m_1/τ fixed. An alternative specialization has a degenerate distribution for ϕ. This occurs if $m \to \infty$ with m/τ fixed. In this case the U_{it} are i.i.d. under P_θ (without conditioning on h_i), and the marginal distribution is a scaled t-distribution with parameter m_1. (This "degrees of freedom" parameter need not be an integer.) In our original model (M), the vector $(U_{i1}, \ldots, U_{iT})$ has a scaled multivariate t-distribution (with parameter m), but the U_{it} are not independent (unconditional on h_i), due to the heterogeneity in volatility. Both of these specializations imply that the marginal distribution of U_{it} is a t-distribution, but we shall see that they have quite different implications for the predictive distributions.

The prior distribution for θ is the same as for our original model, with the addition that m_1 has an independent uniform prior on a discrete grid extending from .1 to 25 with steps of size .01. Table 1.5 reports the medians (and .025 and .975 quantiles) of the posterior distribution based on the high-school sample. The posterior median for m is 1.35, with a .95 probability interval of (.91, 2.01). With our original model, the high-school sample gives a posterior median for m of 1.14 (Table 1.1) and a .95 interval of (.79, 1.58). So heterogeneity in volatility appears to still be important, with no indication that the U_{it} are independent over time (unconditional on h_i). The posterior median for m_1 is 2.32 with a .95 interval of (1.72, 3.28). So ζ_{it} provides an additional source of nonnormality in the cross-sectional distribution of U_{it}.

Figure 1.8 provides the predictive distributions for individuals $i = 297$ and $i = 481$. The predictive distributions in panel (b) are based on the original model (M), in which $h_{it} = \phi_i$. Panel (c) has the other specialization, with $h_{it} = \zeta_{it}/\tau$ and thus no heterogeneity in volatility. Panel (d) is based on the general model (M_1). We see that the original model and the extension (M_1) produce similar predictive distributions, with the spread at horizon $T+1$ being very sensitive to the volatility of the sample history. In contrast, the predictive distributions in panel (c) do not exhibit this sensitivity; they are more like the predictive distributions based on the multivariate normal model in Figure 1.7, in which the spread of the predictive distribution does not depend upon the individual's sample history.

1.5.2 *Heterogeneity in the Volatility of Autoregressive Shocks*

The variance-components model for volatility seems useful, and we would like to apply it to the autoregressive shocks W_{it} in addition to the transitory shocks U_{it}. But first we shall consider simplifying the model. We have repeatedly found that the autoregressive parameter γ is close to one. If $\gamma = 1$, then we cannot distinguish between the individual effect α_i and the initial condition V_{i1} for the random walk, for we can set $\tilde{V}_{i1} = V_{i1} + \alpha_i$, $\tilde{\alpha}_i \equiv 0$, and have an equivalent model. It is plausible that we can drop α_i from the model and obtain similar predictive distributions. With this simplification, our next extension is

$$Y_{it} = V_{i,t} + U_{it} \qquad (t = 1, \ldots, H) \qquad (M_2)$$

$$V_{it} = \gamma V_{i,t-1} + W_{it} \qquad (t = 2, \ldots, H),$$

where

$$V_{i1} \sim \mathcal{N}(0, \sigma_v^2), \quad W_{it} \,|\, \phi_{wi}, \zeta_{wi} \overset{\text{ind}}{\sim} \mathcal{N}(0, h_{wit}^{-1}), \quad U_{it} \,|\, \phi_i, \zeta_i \overset{\text{ind}}{\sim} \mathcal{N}(0, h_{it}^{-1})$$

$$h_{it} = \phi_i \zeta_{it}, \quad \zeta_i = \{\zeta_{it}\}_{t=1}^H, \quad h_{wit} = \phi_{wi}\zeta_{wit}, \quad \zeta_{wi} = \{\zeta_{wit}\}_{t=2}^H$$

$$\phi_i \sim \mathcal{G}(m/2, \tau/2), \quad \zeta_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(m_1/2, 1/2)$$

$$\phi_{wi} \sim \mathcal{G}(m_2/2, \tau_w/2), \quad \zeta_{wit} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(m_3/2, 1/2).$$

Furthermore, V_{i1}, $(\phi_{wi}, \zeta_{wi}, \{W_{it}\}_{t=2}^H)$, and $(\phi_i, \zeta_i, \{U_{it}\}_{t=1}^H)$ are mutually independent under P_θ; also, ϕ_i and ζ_i are independent, and ϕ_{wi} and ζ_{wi} are independent. The parameter vector is $\theta = (\gamma, \sigma_v, m, \tau, m_1, m_2, \tau_w, m_3)$.

The prior distribution for the parameters in the residual process is again chosen to be relatively uninformative. The prior for γ is $\mathcal{N}(0, (.0001)^{-1})$. The priors for m, m_1, m_2, and m_3 are uniform on a discrete grid extending from .1 to 25 with steps of size .01. The priors for σ_v^{-2}, τ, and τ_w are $\chi^2(1)/.01$. All these components of θ are independent under the prior.

Posterior results based on the high-school sample are in Table 1.5. The posterior median for m is 1.29 with a .95 interval of (.81, 2.13). So heterogeneity in volatility still appears to be important for the transitory shocks U_{it}. This is born out in the predictive distributions, which are in Figure 1.9. Panel (a) is based on model (M_1), including α_i. Panel (b) provides a comparison in which we drop α_i. (This corresponds to model (M_2) with degenerate distributions for ϕ_{wi} and ζ_{wit}, so that h_{wit} is constant.) We see that the predictive distributions with and without α_i are very similar. Panel (c) is based on model (M_2). Again we find that the individual with the more volatile earnings history $(i = 297)$

has the predictive distributions that are more spread out. In fact, the contrast between the predictive distributions for $i = 297$ and $i = 481$ is sharper than for model (M_1), reflecting the additional heterogeneity in the volatility of W_{it}. This heterogeneity is captured by m_2, which has a posterior median of 3.17 and a .95 interval of (2.00, 6.44). (The posterior distribution of m_2 is highly skewed.)

Posterior results based on the college sample are in Table 1.6 and Figure 1.10. The results for model (M_1) in columns (1) and (2) are similar to the results for the high school sample in Table 1.5. Figure 1.10 has predictive densities based on model M_1 in panel (a) (with α_i included) and panel (b) (dropping α_i). These predictive densities are similar to the ones in Figure 1.9 for the high school sample, and again dropping α_i makes little difference.

The results for model (M_2) using the college sample are in column 3 of Table 1.6. Now m is poorly determined, so there is not clear evidence for heterogeneity in the volatility of U_{it}. But there is evidence for heterogeneity in the volatility of W_{it}: the posterior median of m_2 is 2.50 with a .95 interval of (1.70, 4.18). The predictive densities for $i = 297$ and $i = 481$ are in Figure 1.10, panel (c). The spreads of these predictive densities are still very sensitive to the differing amounts of volatility in the earnings histories.

1.5.3 *Slope Heterogeneity*

Our final extension combines model (M_1) with heterogeneity in the autoregressive slope γ_i:

$$Y_{it} = V_{it} + U_{it} \qquad (t = 1, \ldots, H) \tag{M_3}$$

$$V_{it} = \gamma_i V_{i,t-1} + W_{it} \qquad (t = 2, \ldots, H),$$

where

$$V_{i1} \sim \mathcal{N}(0, \sigma_v^2), \quad W_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma_w^2), \quad U_{it} \,|\, \phi_i, \zeta_i \overset{\text{ind}}{\sim} \mathcal{N}(0, h_{it}^{-1})$$

$$h_{it} = \phi_i \zeta_{it}, \quad \zeta_i = \{\zeta_{it}\}_{t=1}^H, \quad \phi_i \sim \mathcal{G}(m/2, \tau/2), \quad \zeta_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(m_1/2, 1/2)$$

$$\gamma_i \sim \mathcal{N}(\bar{\gamma}, \sigma_\gamma^2).$$

Furthermore, V_{i1}, $\{W_{it}\}_{t=2}^H$, $(\phi_i, \zeta_i, \{U_{it}\}_{t=1}^H)$, and γ_i are mutually independent under P_θ, and ϕ_i and ζ_i are independent. The parameter vector θ consists of σ_v, σ_w, m, τ, m_1, $\bar{\gamma}$, and σ_γ.

The prior for $\bar{\gamma}$ is $\mathcal{N}(0, (.0001)^{-1})$. The priors for m and m_1 are uniform on the same discrete grid as before. The priors for σ_v^{-2}, σ_w^{-2}, τ, and σ_γ^{-2} are $\chi^2(1)/.01$. All these components of θ are independent under the prior.

Posterior results based on the high school sample are in Table 1.5. The posterior distribution for m is not much affected by allowing for slope heterogeneity. The median is 1.33 with a .95 interval of (.93, 1.98). The predictive distributions for $i = 297$ and $i = 481$ are in Figure 1.9, panel (d). The consequences of heterogeneity in volatility are still apparent, with predictive distributions at $T + 1$ that are much more spread out for $i = 297$ than for $i = 481$. Also note that the predictive distributions for $i = 297$ become extremely dispersed at longer horizons. This reflects the posterior distribution for γ_{297} putting substantial probability above one, so that the predictive distribution for $T + h$ is heavily influenced by the contribution from raising γ to the power h. It might be of interest to consider prior distributions for $\{\gamma_i\}_{i=1}^N$ that restrict the amount of probability placed above one. Results for the college sample are in Table 1.6, column (4) and Figure 1.10, panel (d). These are similar to the results for the high school sample.

1.6 Conclusion

We have focused on using longitudinal data to produce predictive distributions. We have used parametric models, but the parameter estimates have not been of central interest. Rather, the parametric model is a convenient device for producing predictive distributions.

Our main modification of previous models has been to allow for heterogeneity in volatility. This has important consequences, in that the spread of the predictive distribution becomes sensitive to the variability in the earnings history. This conclusion holds up in the various extensions of the model that we have examined.

Appendix

We use Gibbs sampling, as in Gelfand and Smith (1990), with parameter augmentation as in Tanner and Wong (1987). Gilks, Richardson, and Spiegelhalter (1996) provide a review of these methods with references to the recent literature. This appendix sketches our algorithms by showing the blocks of parameters and how we sample at each block, conditional on the values for the parameters in the other blocks. We start with Model (M_1), which specializes to model (M). Then we present the algorithms for models (M_2) and (M_3).

A.1 *Model (M_1)*

Blocks: ϕ_i $(i = 1, \ldots, N)$, ζ_{it} $(i = 1, \ldots, N; t = 1, \ldots, T)$, σ_v, (γ, σ_w), (ψ, σ_ϵ), (v_i, α_i) $(i = 1, \ldots, N)$, (m, τ), m_1. (v_i is defined as $v_i = (v_{i1}, \ldots, v_{iT})'$.)

ϕ_i:

$$\phi_i \sim \mathcal{G}(\frac{T+m}{2}, \frac{1}{2})/[\sum_{t=1}^{T} \zeta_{it}(y_{it} - v_{it} - \alpha_i)^2 + \tau].$$

ζ_{it}:

$$\zeta_{it} \sim \mathcal{G}(\frac{1+m_1}{2}, \frac{1}{2})/[\phi_i(y_{it} - v_{it} - \alpha_i)^2 + 1].$$

σ_v:

$$\sigma_v^{-2} \sim \mathcal{G}(\frac{N+1}{2}, \frac{1}{2})/[\sum_{i=1}^{N} v_{i1}^2 + .01].$$

(γ, σ_w):

$$\sigma_w^{-2} \sim \mathcal{G}(\frac{N(T-1)-1}{2}, \frac{1}{2})/\sum_{i=1}^{N}\sum_{t=2}^{T}(v_{it} - \hat{\gamma} v_{i,t-1})^2$$

$$\gamma \,|\, \sigma_w \;\sim\; \mathcal{N}(\hat{\gamma}, \sigma_w^2 (\sum_{i=1}^{N} \sum_{t=2}^{T} v_{i,t-1}^2)^{-1}),$$

where

$$\hat{\gamma} = (\sum_{i=1}^{N} \sum_{t=2}^{T} v_{i,t-1}^2)^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} v_{i,t-1} v_{it}.$$

(ψ, σ_ϵ):

$$\sigma_\epsilon^{-2} \;\sim\; \mathcal{G}(\frac{N}{2}, \frac{1}{2})/[\sum_{i=1}^{N} (\alpha_i - \hat{\psi} v_{i1})^2 + .01],$$

$$\psi \,|\, \sigma_\epsilon \;\sim\; \mathcal{N}(\hat{\psi}, \sigma_\epsilon^2 (\sum_{i=1}^{N} v_{i1}^2)^{-1}),$$

where

$$\hat{\psi} = (\sum_{i=1}^{N} v_{i1}^2)^{-1} \sum_{i=1}^{N} v_{i1} \alpha_i.$$

(v_i, α_i):

$$v_i \;\sim\; \mathcal{N}((H + H_1)^{-1} F' \Omega^{-1} y_i, (H + H_1)^{-1}),$$

where

$$H = F' \Omega^{-1} F, \quad H_1 = B' \begin{pmatrix} \sigma_v^{-2} & 0 \\ 0 & \sigma_w^{-2} I_{T-1} \end{pmatrix} B,$$

and

$$F = \begin{pmatrix} \psi + 1 & 0 & \dots & 0 \\ \psi & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \psi & 0 & \dots & 1 \end{pmatrix}, \quad y_i = \begin{pmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{pmatrix}, \quad h_i = \begin{pmatrix} h_{i1} \\ \vdots \\ h_{iT} \end{pmatrix}, \quad l = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix},$$

$$\Omega^{-1} = (\sigma_\epsilon^2 l l' + \mathrm{diag}\{h_{i1}^{-1}, \dots, h_{iT}^{-1}\})^{-1} = \mathrm{diag}\{h_{i1}, \dots, h_{iT}\} - \sigma_\epsilon^2 h_i h_i' / (1 + \sigma_\epsilon^2 \sum_{t=1}^{T} h_{it}),$$

$$B = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ -\gamma & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & -\gamma & 1 \end{pmatrix}.$$

$$\epsilon_i \,|\, v_i \sim \mathcal{N}\big((\sum_{t=1}^{T} h_{it} + \sigma_\epsilon^{-2})^{-1} \sum_{t=1}^{T} h_{it}(y_{it} - v_{it} - \psi v_{i1}), (\sum_{t=1}^{T} h_{it} + \sigma_\epsilon^{-2})^{-1}\big)$$

$$\alpha_i = \psi v_{i1} + \epsilon_i.$$

(m, τ): The conditional density for (m, τ) (given the data and the parameter values for the other blocks) is

$$c\Big\{[\frac{1}{2}(\sum_{i=1}^{N} \phi_i + .01)]^{\frac{mN+1}{2}} \tau^{[\frac{mN+1}{2}-1]} \exp(-\tau[\frac{1}{2}(\sum_{i=1}^{N} \phi_i + .01)])/\Gamma(\frac{mN+1}{2})\Big\}$$
$$\times \Big\{(\prod_{i=1}^{N} \phi_i)^{\frac{m}{2}-1} \Gamma(\frac{mN+1}{2}) \Big/ \Big[[\frac{1}{2}(\sum_{i=1}^{N} \phi_i + .01)]^{\frac{mN+1}{2}} [\Gamma(\frac{m}{2})]^N\Big]\Big\},$$

where c is a normalizing constant that does not depend upon m or τ. This density is with respect to the product of counting measure on $\{.01, .02, ..., 25\}$ and Lebesgue measure on the positive real line. We recognize the first term in braces as a gamma density. So it is the conditional density for τ given m, and the second term in braces is proportional to the marginal density for m. Hence the sampling procedure is as follows:

$$m \sim \sum_{j=1}^{J} p_j \delta_{a_j},$$

where δ_a is a unit point mass at a, $J = 2491$, $a_j = .09 + (.01)j$,

$$\log \tilde{p}_j = (\frac{a_j}{2} - 1) \sum_{i=1}^{N} \log \phi_i + \log \Gamma(\frac{a_j N + 1}{2})$$
$$- \frac{a_j N + 1}{2} \log(\sum_{i=1}^{N} \phi_i + .01) - N \log \Gamma(a_j/2)$$

$$p_j = \tilde{p}_j / \sum_{j=1}^{J} \tilde{p}_j.$$

$$\tau \,|\, m \sim \mathcal{G}(\frac{mN+1}{2}, \frac{1}{2})/(\sum_{i=1}^{N} \phi_i + .01).$$

m_1:

$$m_1 \sim \sum_{j=1}^{J} p_{1j}\delta_{a_j},$$

where

$$\log \tilde{p}_{1j} = \frac{a_j}{2}\sum_{i=1}^{N}\sum_{t=1}^{T}\log(\zeta_{it}/2) - NT\log\Gamma(a_j/2),$$

$$p_{1j} = \tilde{p}_{1j}/\sum_{j=1}^{J}\tilde{p}_{1j}.$$

Predictive Distribution

Let $i = d \in \{1, \ldots, N\}$ denote the decision maker.

$$(Y_{d1}, \ldots, Y_{dH}) \,|\, \lambda_d \sim \mathcal{N}(0, \Sigma(\lambda_d)),$$

where $\lambda_d = (\theta, \phi_d, \zeta_{d1}, \ldots, \zeta_{dH})$, $\theta = (\gamma, \psi, \sigma_v, \sigma_w, \sigma_\epsilon, m, \tau, m_1)$, and

$$\Sigma(\lambda) = A(\gamma,\psi)[\mathrm{diag}\{\sigma_v^2, \sigma_w^2, \ldots, \sigma_w^2\}]A(\gamma,\psi)' + \sigma_\epsilon^2 ll' + \mathrm{diag}\{h_1^{-1}, \ldots, h_H^{-1}\},$$

where

$$A(\gamma,\psi) = \begin{pmatrix} 1+\psi & 0 & 0 & \ldots & 0 \\ \gamma+\psi & 1 & 0 & \ldots & 0 \\ \gamma^2+\psi & \gamma & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma^{H-1}+\psi & \gamma^{H-2} & \gamma^{H-3} & \ldots & 1 \end{pmatrix},$$

l is a $H \times 1$ vector of ones, and $h_t = \phi\zeta_t$ $(t = 1, \ldots, H)$. Hence

$$(Y_{d,T+1}, \ldots, Y_{dH}) \,|\, y_d^T, \lambda_d \sim \mathcal{N}(\mu(\lambda_d), \Omega(\lambda_d)),$$

where

$$\mu(\lambda) = (\Sigma_{11}^{-1}(\lambda)\Sigma_{12}(\lambda))' y_d^T, \tag{A.1}$$

$$\Omega(\lambda) = \Sigma_{22}(\lambda) - \Sigma_{21}(\lambda)\Sigma_{11}^{-1}(\lambda)\Sigma_{12}(\lambda),$$

and we have partitioned the $H \times H$ matrix $\Sigma(\lambda)$ as

$$\Sigma(\lambda) = \begin{pmatrix} \Sigma_{11}(\lambda) & \Sigma_{12}(\lambda) \\ \Sigma_{21}(\lambda) & \Sigma_{22}(\lambda) \end{pmatrix},$$

where $\Sigma_{11}(\lambda)$ is $T \times T$.

Let $\tilde{\lambda}_d^{(j)} = (\theta^{(j)}, \phi_d^{(j)}, \zeta_{d1}^{(j)}, \ldots, \zeta_{dT}^{(j)})$ denote the j^{th} draw from the posterior distribution of $(\theta, \phi_d, \zeta_{d1}, \ldots, \zeta_{dT})$ (from Gibbs sampling, after some burn-in period). Let

$$\zeta_{ds}^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(\frac{m_1^{(j)}}{2}, \frac{1}{2}) \qquad (s = T+1, \ldots, H),$$

and set $\lambda_d^{(j)} = (\tilde{\lambda}_d^{(j}, \zeta_{d,T+1}^{(j)}, \ldots, \zeta_{dH}^{(j)})$. Let $\Phi(\cdot \,|\, \mu, \Omega)$ denote the $(H-T)$-variate normal distribution function, with mean μ and covariance matrix Ω. Let $F(\cdot \,|\, y_d^T, z)$ denote the distribution function for the predictive distribution of $(Y_{d,T+1}, \ldots, Y_{dH})$ conditional on $Y_d^T = y_d^T$ and on the sample data z, corresponding to the density in equation (1). Then our Monte Carlo evaluation of the predictive distribution is given by

$$F(y_{T+1}, \ldots, y_H \,|\, y_d^T, z) \;\approx\; \frac{1}{M} \sum_{j=1}^{M} \Phi(y_{T+1}, \ldots, y_H \,|\, \mu(\lambda^{(j)}), \Omega(\lambda^{(j)})). \tag{A.2}$$

A.2 *Model* (M_2)

Blocks: ϕ_i $(i = 1, \ldots, N)$, ζ_{it} $(i = 1, \ldots, N;\, t = 1, \ldots, T)$, σ_v, ϕ_{wi} $(i = 1, \ldots, N)$, ζ_{wit} $(i = 1, \ldots, N;\, t = 2, \ldots, T)$, γ, v_i $(i = 1, \ldots, N)$, (m, τ), m_1, (m_2, τ_w), m_3.

ϕ_i, ζ_{it}, σ_v: as in Model (M_1), dropping α_i.

ϕ_{wi}:

$$\phi_{wi} \sim \mathcal{G}(\frac{T-1+m_2}{2}, \frac{1}{2})/[\sum_{t=2}^{T} \zeta_{wit}(v_{it} - \gamma v_{i,t-1})^2 + \tau_w]$$

ζ_{wit}:

$$\zeta_{wit} \sim \mathcal{G}(\frac{1+m_3}{2}, \frac{1}{2})/[\phi_{wi}(v_{it} - \gamma v_{i,t-1})^2 + 1].$$

γ:

$$\gamma \sim \mathcal{N}((H_\gamma + .0001)^{-1} \sum_{i=1}^{N} \sum_{t=2}^{T} h_{wit} v_{i,t-1} v_{it}, (H_\gamma + .0001)^{-1}),$$

where

$$H_\gamma = \sum_{i=1}^{N} \sum_{t=2}^{T} h_{wit} v_{i,t-1}^2.$$

v_i:

$$v_i \sim \mathcal{N}((H + H_1)^{-1} H y_i, (H + H_1)^{-1}),$$

where

$$H = \mathrm{diag}\{h_{i1}, \ldots, h_{iT}\}, \quad H_1 = B' \mathrm{diag}\{\sigma_v^{-2}, h_{wi2}, \ldots, h_{wiT}\} B,$$

and B is defined as above.

(m, τ), m_1: as in model (M_1).

(m_2, τ_w):

$$m_2 \sim \sum_{j=1}^{J} p_{2j} \delta_{a_j},$$

where δ_a is a point mass at a, $J = 2491$, $a_j = .09 + (.01)j$,

$$\log \tilde{p}_{2j} = (\frac{a_j}{2} - 1) \sum_{i=1}^{N} \log \phi_{wi} + \log \Gamma(\frac{a_j N + 1}{2})$$
$$- \frac{a_j N + 1}{2} \log(\sum_{i=1}^{N} \phi_{wi} + .01) - N \log \Gamma(a_j/2),$$
$$p_{2j} = \tilde{p}_{2j} / \sum_{j=1}^{J} \tilde{p}_{2j}.$$

$$\tau_w \,|\, m_2 \;\sim\; \mathcal{G}(\frac{m_2 N + 1}{2}, \frac{1}{2})/(\sum_{i=1}^{N} \phi_{wi} + .01).$$

m_3:

$$m_3 \;\sim\; \sum_{j=1}^{J} p_{3j} \delta_{a_j},$$

where

$$\log \tilde{p}_{3j} = \frac{a_j}{2} \sum_{i=1}^{N} \sum_{t=2}^{T} \log(\zeta_{wit}/2) - N(T-1) \log \Gamma(a_j/2),$$

$$p_{3j} = \tilde{p}_{3j} / \sum_{j=1}^{J} \tilde{p}_{3j}.$$

Predictive Distribution

$$(Y_{d1}, \ldots, Y_{dH}) \,|\, \lambda_d \;\sim\; \mathcal{N}(0, \Sigma(\lambda_d)),$$

where $\lambda_d = (\theta, \phi_d, \{\zeta_{dt}\}_{t=1}^{H}, \phi_{wd}, \{\zeta_{wdt}\}_{t=2}^{H})$, $\theta = (\gamma, \sigma_v, m, \tau, m_1, m_2, \tau_w, m_3)$, and

$$\Sigma(\lambda) = A(\gamma)[\mathrm{diag}\{\sigma_v^2, h_{w2}^{-1}, \ldots, h_{wH}^{-1}\}]A(\gamma)' + \mathrm{diag}\{h_1^{-1}, \ldots, h_H^{-1}\},$$

where

$$A(\gamma) = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ \gamma & 1 & 0 & \ldots & 0 \\ \gamma^2 & \gamma & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \gamma^{H-1} & \gamma^{H-2} & \gamma^{H-3} & \ldots & 1 \end{pmatrix}, \qquad \text{(A.3)}$$

$h_t = \phi \zeta_t$, and $h_{wt} = \phi_w \zeta_{wt}$. Hence

$$(Y_{d,T+1}, \ldots, Y_{dH}) \,|\, y_d^T, \lambda_d \;\sim\; \mathcal{N}(\mu(\lambda_d), \Omega(\lambda_d)),$$

where μ and Ω are defined in (A.1).

Let $(\theta^{(j)}, \phi_d^{(j)}, \{\zeta_{dt}^{(j)}\}_{t=1}^T, \phi_{wd}^{(j)}, \{\zeta_{wdt}^{(j)}\}_{t=2}^T)$ denote the j^{th} draw from the posterior distribution of $(\theta, \phi_d, \{\zeta_{dt}\}_{t=1}^T, \phi_{wd}, \{\zeta_{wdt}\}_{t=2}^T)$ (from Gibbs sampling). Let

$$\zeta_{ds}^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(\frac{m_1^{(j)}}{2}, \frac{1}{2}), \quad \zeta_{wds}^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(\frac{m_3^{(j)}}{2}, \frac{1}{2}) \qquad (s = T+1, \ldots, H),$$

and set $\lambda_d^{(j)} = (\theta^{(j)}, \phi_d^{(j)}, \{\zeta_{dt}^{(j)}\}_{t=1}^H, \phi_{wd}^{(j)}, \{\zeta_{wdt}^{(j)}\}_{t=2}^H)$. Then, as in (A.2), our Monte Carlo evaluation of the predictive distribution is given by

$$F(y_{T+1}, \ldots, y_H \,|\, y_d^T, z) \;\approx\; \frac{1}{M} \sum_{j=1}^{M} \Phi(y_{T+1}, \ldots, y_H \,|\, \mu(\lambda^{(j)}), \Omega(\lambda^{(j)})).$$

A.3 *Model* (M_3)

Blocks: ϕ_i $(i = 1, \ldots, N)$, ζ_{it} $(i = 1, \ldots, N;\, t = 1, \ldots, T)$, σ_v, σ_w, γ_i $(i = 1, \ldots, N)$, $\bar{\gamma}$, σ_γ, v_i $(i = 1, \ldots, N)$, (m, τ), m_1.

ϕ_i, ζ_{it}, σ_v: as in model (M_1), dropping α_i.

σ_w:

$$\sigma_w^{-2} \;\sim\; \mathcal{G}(\frac{N(T-1)+1}{2}, \frac{1}{2}) / \sum_{i=1}^{N} \sum_{t=2}^{T} ((v_{it} - \gamma_i v_{i,t-1})^2 + .01).$$

γ_i:

$$\gamma_i \;\sim\; \mathcal{N}((H_\gamma + \sigma_\gamma^{-2})^{-1}(H_\gamma \hat{\gamma}_i + \sigma_\gamma^{-2}\bar{\gamma}), (H_\gamma + \sigma_\gamma^{-2})^{-1})$$

where

$$H_\gamma = \sigma_w^{-2} \sum_{t=2}^{T} v_{i,t-1}^2, \quad \hat{\gamma}_i = (\sum_{t=2}^{T} v_{i,t-1}^2)^{-1} \sum_{t=2}^{T} v_{i,t-1} v_{it}.$$

$\bar{\gamma}$:

$$\bar{\gamma} \;\sim\; \mathcal{N}((\sigma_\gamma^{-2} N + .0001)^{-1} \sigma_\gamma^{-2} \sum_{i=1}^{N} \gamma_i, (\sigma_\gamma^{-2} N + .0001)^{-1}).$$

σ_γ:

$$\sigma_\gamma^{-2} \sim \mathcal{G}(\frac{N+1}{2}, \frac{1}{2}) \Big/ (\sum_{i=1}^{N}(\gamma_i - \bar{\gamma})^2 + .01).$$

v_i:

$$v_i \sim \mathcal{N}((H+H_1)^{-1}Hy_i, (H+H_1)^{-1}),$$

where

$$H = \text{diag}\{h_{i1}, \ldots, h_{iT}\}, \quad H_1 = B'\text{diag}\{\sigma_v^{-2}, \sigma_w^{-2}, \ldots, \sigma_w^{-2}\}B,$$

$$B = \begin{pmatrix} 1 & 0 & \ldots & 0 & 0 \\ -\gamma_i & 1 & \ldots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \ldots & -\gamma_i & 1 \end{pmatrix}.$$

(m, τ), m_1: as in model (M_1).

Predictive Distribution

$$(Y_{d1}, \ldots, Y_{dH}) \,|\, \lambda_d \sim \mathcal{N}(0, \Sigma(\lambda_d)),$$

where $\lambda_d = (\theta, \phi_d, \{\zeta_{dt}\}_{t=1}^{H}, \gamma_d)$, $\theta = (\sigma_v, \sigma_w, m, \tau, m_1, \bar{\gamma}, \sigma_\gamma)$, and

$$\Sigma(\lambda) = A(\gamma)[\text{diag}\{\sigma_v^2, \sigma_w^2, \ldots, \sigma_w^2\}]A(\gamma)' + \text{diag}\{h_1^{-1}, \ldots, h_H^{-1}\},$$

where $A(\gamma)$ is defined in (A.3) and $h_t = \phi\zeta_t$ $(t = 1, \ldots, H)$. Hence

$$(Y_{d,T+1}, \ldots, Y_{dH}) \,|\, y_d^T, \lambda_d \sim \mathcal{N}(\mu(\lambda_d), \Omega(\lambda_d)),$$

where μ and Ω are defined in (A.1).

Let $(\theta^{(j)}, \phi_d^{(j)}, \{\zeta_{dt}^{(j)}\}_{t=1}^{T}, \gamma_d)$ denote the j^{th} draw from the posterior distribution of $(\theta, \phi_d, \{\zeta_{dt}\}_{t=1}^{T}, \gamma_d)$ (from Gibbs sampling). Let

$$\zeta_{ds}^{(j)} \overset{\text{i.i.d.}}{\sim} \mathcal{G}(\frac{m_1^{(j)}}{2}, \frac{1}{2}) \qquad (s = T+1, \ldots, H),$$

and set $\lambda_d^{(j)} = (\theta^{(j)}, \phi_d^{(j)}, \{\zeta_{dt}^{(j)}\}_{t=1}^{H}, \gamma_d^{(j)})$. Then, as in (A.2), our Monte Carlo evaluation of the predictive distribution is given by

$$F(y_{T+1}, \ldots, y_H \,|\, y_d^T, z) \;\approx\; \frac{1}{M} \sum_{j=1}^{M} \Phi(y_{T+1}, \ldots, y_H \,|\, \mu(\lambda^{(j)}), \Omega(\lambda^{(j)})).$$

TABLE 1.1

POSTERIOR FOR θ

Parameter	Full Sample $N = 813$	H.S. Dropouts $N = 37$	H.S. Grads $N = 100$	Col. Grads $N = 122$
γ	.98 $(.97, .99)$	.95 $(.72, 1.05)$	.97 $(.89, 1.02)$	.99 $(.94, 1.06)$
ψ	$-.00$ $(-.11, .09)$	$-.12$ $(-2.92, 2.18)$	.00 $(-2.21, 2.17)$	$-.06$ $(-2.81, 1.74)$
σ_v	.32 $(.23, .35)$	.12 $(.04, .34)$	.15 $(.05, .38)$	.15 $(.04, .38)$
σ_w	.10 $(.10, .11)$	.11 $(.08, .15)$,	.11 $(.09, .13)$	.10 $(.08, .11)$
σ_ϵ	.08 $(.04, .22)$	.23 $(.08, .33)$	.27 $(.07, .36)$	.26 $(.06, .35)$
m	1.20 $(1.07, 1.33)$	1.63 $(.94, 2.78)$	1.14 $(.79, 1.58)$	1.14 $(.85, 1.52)$
$\bar{\sigma}_u$	.11 $(.10, .12)$	.16 $(.11, .21)$	.11 $(.08, .13)$	.09 $(.07, .11)$

Note: .025 and .975 quantiles are in parentheses.

TABLE 1.2

PREDICTIVE DISTRIBUTION FOR $\sigma_u = 1/\sqrt{h}$

	Quantile						
Sample	.05	.10	.25	.50	.75	.90	.95
Full ($N = 813$)	.04	.05	.07	.11	.20	.45	.81
H.S. Dropouts ($N = 37$)	.06	.08	.11	.16	.27	.49	.79
H.S. Grads ($N = 100$)	.04	.05	.06	.11	.21	.49	.93
Col. Grads ($N = 122$)	.03	.04	.05	.09	.18	.40	.75

TABLE 1.3

PREDICTIVE DISTRIBUTIONS FOR log(EARNINGS) RESIDUALS

	Quantile						
Individual	.05	.10	.25	.50	.75	.90	.95
$i = 38$							
$T+1$	-1.21	-1.06	$-.83$	$-.58$	$-.34$	$-.10$	.05
$T+5$	-1.25	-1.08	$-.82$	$-.54$	$-.26$	.00	.16
$T+10$	-1.28	-1.10	$-.81$	$-.50$	$-.19$	.10	.28
$T+20$	-1.30	-1.10	$-.78$	$-.42$	$-.07$	.25	.45
$\bar{y}_i = -.66,\ y_{i,T} = -.11,\ \mathrm{std}(y_i) = .37$							
$i = 346$							
$T+1$	$-.07$	.05	.23	.42	.62	.81	.94
$T+5$	$-.19$	$-.06$	.16	.40	.64	.86	1.00
$T+10$	$-.31$	$-.15$	.09	.37	.64	.90	1.05
$T+20$	$-.47$	$-.30$	$-.01$	.31	.64	.93	1.11
$\bar{y}_i = .51,\ y_{i,T} = .14,\ \mathrm{std}(y_i) = .28$							
$i = 196$							
$T+1$	$-.82$	$-.71$	$-.54$	$-.35$	$-.14$	.06	.20
$T+5$	$-.89$	$-.77$	$-.56$	$-.32$	$-.08$	.15	.30
$T+10$	$-.96$	$-.81$	$-.56$	$-.29$	$-.01$	.24	.41
$T+20$	-1.03	$-.85$	$-.56$	$-.24$	.08	.38	.56
$\bar{y}_i = -.11,\ y_{i,T} = -.93,\ \mathrm{std}(y_i) = .30$							
$i = 491$							
$T+1$	.14	.23	.36	.49	.60	.70	.76
$T+5$	$-.02$	.09	.27	.45	.63	.78	.88
$T+10$	$-.16$	$-.03$	.18	.41	.63	.84	.96
$T+20$	$-.35$	$-.20$	.06	.34	.63	.88	1.03
$\bar{y}_i = .12,\ y_{i,T} = .65,\ \mathrm{std}(y_i) = .22$							

TABLE 1.3 (continued)

	Quantile						
Individual	.05	.10	.25	.50	.75	.90	.95
$i = 321$							
$T + 1$	$-.29$	$-.25$	$-.17$	$-.09$	$-.00$	.07	.12
$T + 5$	$-.47$	$-.38$	$-.24$	$-.08$	.08	.22	.30
$T + 10$	$-.58$	$-.47$	$-.28$	$-.07$	.14	.32	.44
$T + 20$	$-.72$	$-.57$	$-.33$	$-.06$	.21	.45	.60
$\bar{y}_i = .03,\ y_{i,T} = -.10,\ \mathrm{std}(y_i) = .07$							
$i = 415$							
$T + 1$	$-.94$	$-.76$	$-.48$	$-.18$	.13	.43	.63
$T + 5$	$-.99$	$-.80$	$-.50$	$-.17$	.17	.49	.70
$T + 10$	-1.04	$-.84$	$-.51$	$-.15$	.21	.55	.77
$T + 20$	-1.10	$-.88$	$-.52$	$-.12$	.28	.64	.87
$\bar{y}_i = -.01,\ y_{i,T} = -.48,\ \mathrm{std}(y_i) = .47$							
$i = 481$							
$T + 1$	$-.15$	$-.10$	$-.02$	.06	.15	.23	.28
$T + 5$	$-.33$	$-.24$	$-.10$	.06	.22	.37	.45
$T + 10$	$-.46$	$-.34$	$-.15$	.06	.27	.46	.57
$T + 20$	$-.61$	$-.46$	$-.22$	.05	.32	.57	.71
$\bar{y}_i = .12,\ y_{i,T} = .05,\ \mathrm{std}(y_i) = .07$							
$i = 297$							
$T + 1$	-1.90	-1.67	-1.30	$-.91$	$-.49$	$-.07$	.21
$T + 5$	-1.89	-1.65	-1.26	$-.85$	$-.41$	.02	.31
$T + 10$	-1.88	-1.62	-1.22	$-.78$	$-.32$	.12	.41
$T + 20$	-1.84	-1.57	-1.13	$-.66$	$-.18$	.28	.58
$\bar{y}_i = -1.07,\ y_{i,T} = -.35,\ \mathrm{std}(y_i) = .56$							

TABLE 1.4

PREDICTIVE DISTRIBUTIONS FOR log(EARNINGS) RESIDUALS
RESTRICTED MODEL (h_i constant)

	Quantile						
Individual	.05	.10	.25	.50	.75	.90	.95
$i = 38$							
$T+1$	−.86	−.76	−.58	−.38	−.18	.00	.10
$T+5$	−1.05	−.91	−.69	−.43	−.18	.05	.18
$T+10$	−1.09	−.95	−.71	−.45	−.19	.06	.20
$T+20$	−1.10	−.96	−.72	−.45	−.19	.06	.21
$\bar{y}_i = -.66,\ y_{i,T} = -.11,\ \mathrm{std}(y_i) = .37$							
$i = 346$							
$T+1$	−.16	−.05	.13	.33	.52	.70	.81
$T+5$	−.26	−.12	.11	.36	.61	.83	.97
$T+10$	−.28	−.14	.10	.37	.63	.86	1.00
$T+20$	−.29	−.14	.10	.37	.63	.87	1.01
$\bar{y}_i = .51,\ y_{i,T} = .14,\ \mathrm{std}(y_i) = .28$							
$i = 481$							
$T+1$	−.41	−.30	−.12	.08	.27	.45	.56
$T+5$	−.53	−.39	−.17	.08	.33	.56	.69
$T+10$	−.56	−.41	−.18	.08	.35	.58	.72
$T+20$	−.56	−.42	−.18	.09	.35	.59	.73
$\bar{y}_i = .12,\ y_{i,T} = .05,\ \mathrm{std}(y_i) = .07$							
$i = 297$							
$T+1$	−1.20	−1.10	−.92	−.72	−.52	−.35	−.24
$T+5$	−1.36	−1.23	−1.00	−.75	−.50	−.27	−.13
$T+10$	−1.39	−1.25	−1.02	−.75	−.49	−.25	−.10
$T+20$	−1.40	−1.26	−1.02	−.76	−.49	−.24	−.09
$\bar{y}_i = -1.07,\ y_{i,T} = -.35,\ \mathrm{std}(y_i) = .56$							

TABLE 1.5

POSTERIOR FOR θ—HIGH-SCHOOL SAMPLE

Parameter	(1)	(2)	(3)	(4)
γ	.97 (.92, 1.02)	.98 (.95, 1.01)	.95 (.92, .98)	–
ψ	.93 (−7.39, 6.54)	–	–	–
σ_v	.09 (.04, .37)	.32 (.28, .38)	.34 (.29, .39)	.33 (.28, .38)
σ_w	.11 (.10, .13)	.11 (.10, .13)	–	.11 (.10, .13)
σ_ϵ	.13 (.04, .33)	–	–	–
m	1.35 (.91, 2.01)	1.36 (.90, 2.00)	1.29 (.81, 2.13)	1.33 (.93, 1.98)
m_1	2.32 (1.72, 3.28)	2.31 (1.73, 3.27)	1.87 (1.33, 2.71)	2.32 (1.76, 3.30)
$\bar{\sigma}_u$	.09 (.06, .11)	.09 (.06, .11)	.05 (.03, .07)	.09 (.06, .11)
m_2	–	–	3.17 (2.00, 6.44)	–
m_3	–	–	3.95 (2.50, 7.60)	–
$\bar{\sigma}_w$	–	–	.11 (.09, .14)	–
$\bar{\gamma}$	–	–	–	.96 (.92, .99)
σ_γ	–	–	–	.06 (.03, .10)

Note. (1) Model (M_1): $h_{it} = \phi_i \zeta_{it}$ (include α_i); (2) Model (M_1): $h_{it} = \phi_i \zeta_{it}$ (drop α_i); (3) Model (M_2): $h_{it} = \phi_i \zeta_{it}$, $h_{wit} = \phi_{wit} \zeta_{wit}$; (4) Model ($M_3$): $h_{it} = \phi_i \zeta_{it}$, γ_i; .025 and .975 quantiles are in parentheses.

TABLE 1.6

POSTERIOR FOR θ—COLLEGE SAMPLE

Parameter	(1)	(2)	(3)	(4)
γ	1.05 $(1.01, 1.08)$	.98 $(.96, 1.00)$	.94 $(.92, .97)$	–
ψ	-3.04 $(-7.31, -1.27)$	–	–	–
σ_v	.14 $(.05, .56)$	.33 $(.28, .38)$	.34 $(.29, .39)$	.33 $(.29, .39)$
σ_w	.09 $(.08, .10)$	.10 $(.09, .11)$	–	.11 $(.10, .13)$
σ_ϵ	.13 $(.04, .33)$	–	–	–
m	1.38 $(.97, 1.92)$	1.27 $(.89, 1.76)$	2.52 $(1.16, 23.06)$	1.32 $(.91, 1.86)$
m_1	2.02 $(1.58, 2.69)$	1.97 $(1.53, 2.64)$	1.42 $(1.04, 2.10)$	1.98 $(1.54, 2.67)$
$\bar{\sigma}_u$	.07 $(.05, .09)$	.06 $(.05, .08)$	.03 $(.02, .05)$	.07 $(.05, .08)$
m_2	–	–	2.50 $(1.70, 4.18)$	–
m_3	–	–	3.76 $(2.53, 6.43)$	–
$\bar{\sigma}_w$	–	–	.10 $(.08, .12)$	–
$\bar{\gamma}$	–	–	–	.97 $(.96, .99)$
σ_γ	–	–	–	.07 $(.04, .10)$

Note. (1) Model (M_1): $h_{it} = \phi_i \zeta_{it}$ (include α_i); (2) Model (M_1): $h_{it} = \phi_i \zeta_{it}$ (drop α_i); (3) Model (M_2): $h_{it} = \phi_i \zeta_{it}$, $h_{wit} = \phi_{wit} \zeta_{wit}$; (4) Model ($M_3$): $h_{it} = \phi_i \zeta_{it}$, γ_i; .025 and .975 quantiles are in parentheses.

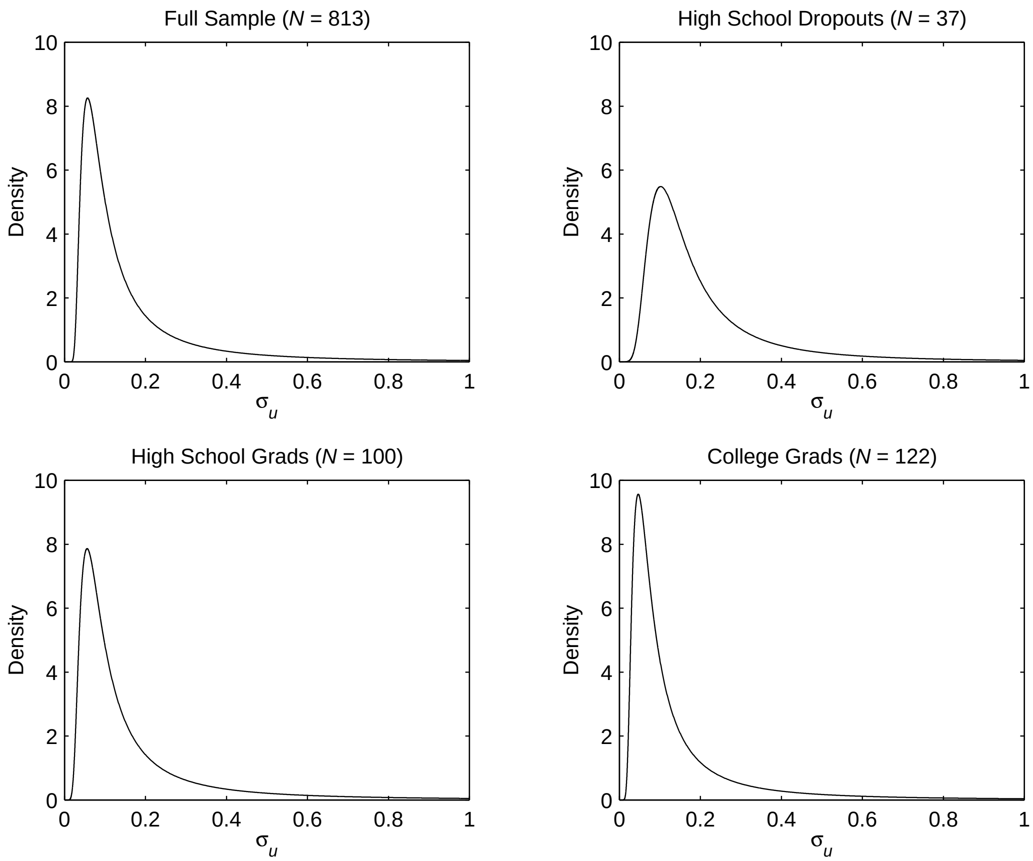

FIGURE 1.1. Predictive densities for $\sigma_u = 1/\sqrt{h}$.

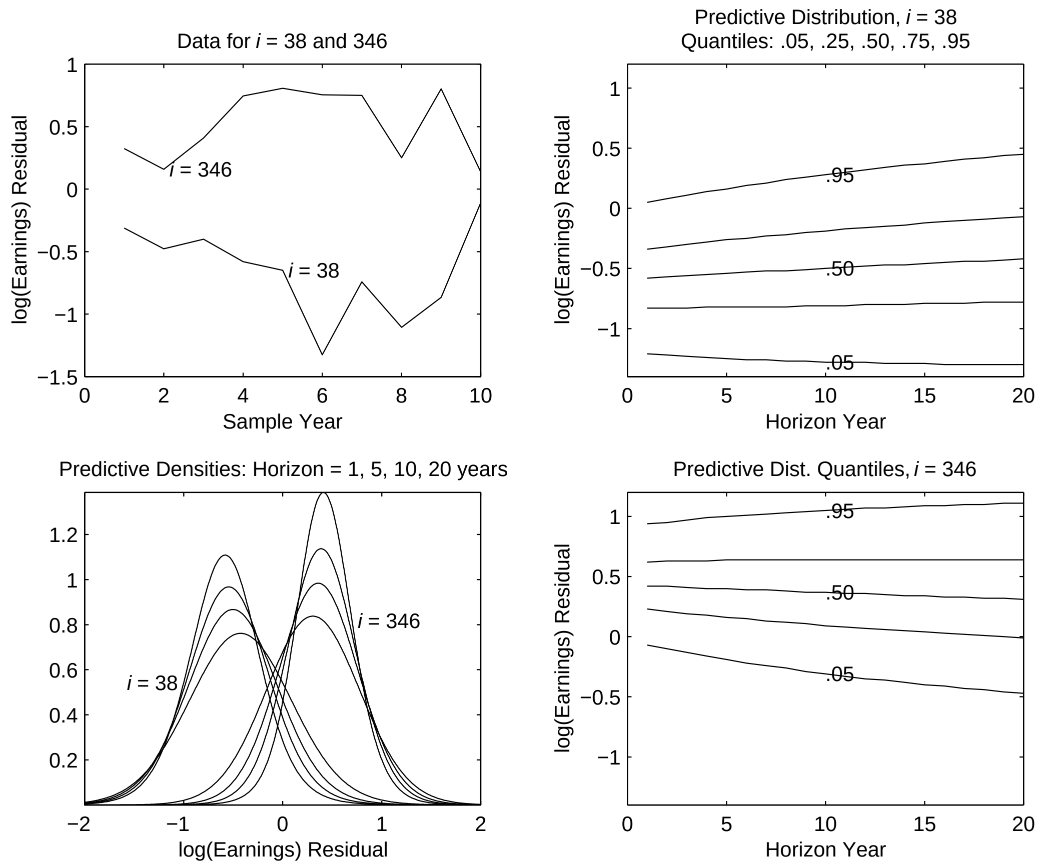

FIGURE 1.2. Data and predictive distributions for individuals $i = 38$ and 346.

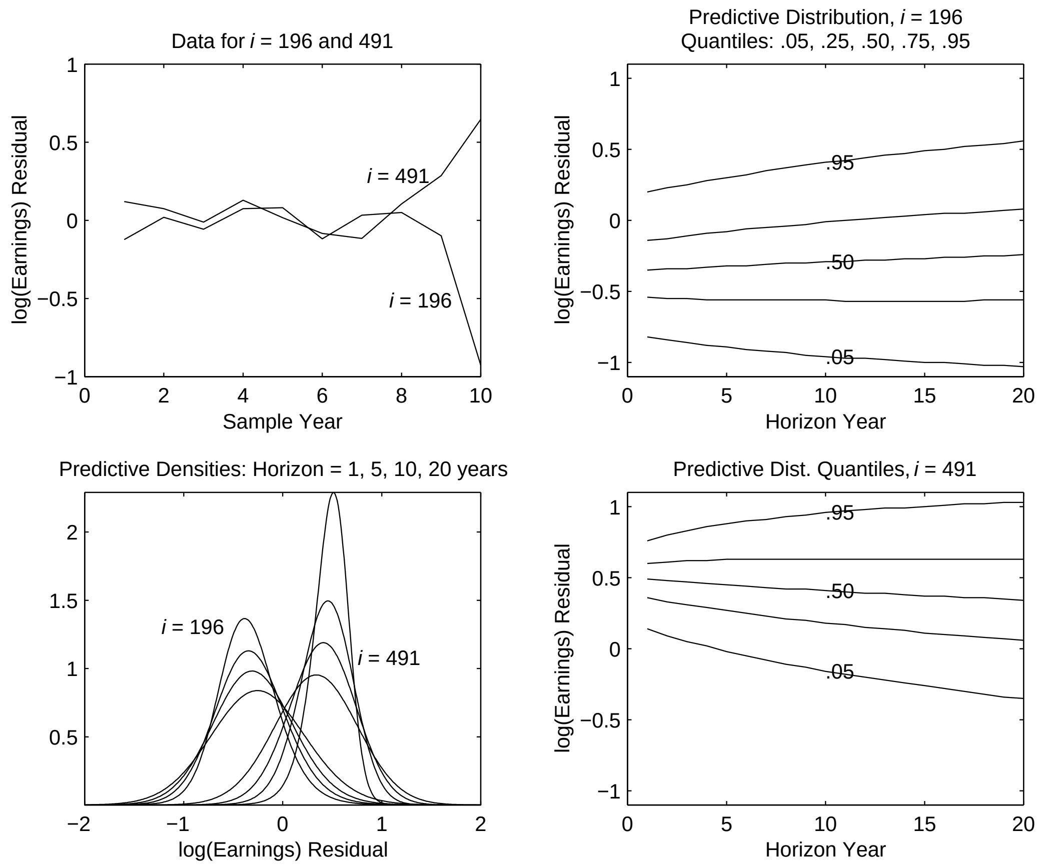

FIGURE 1.3. Data and predictive distributions for individuals $i = 196$ and 491.

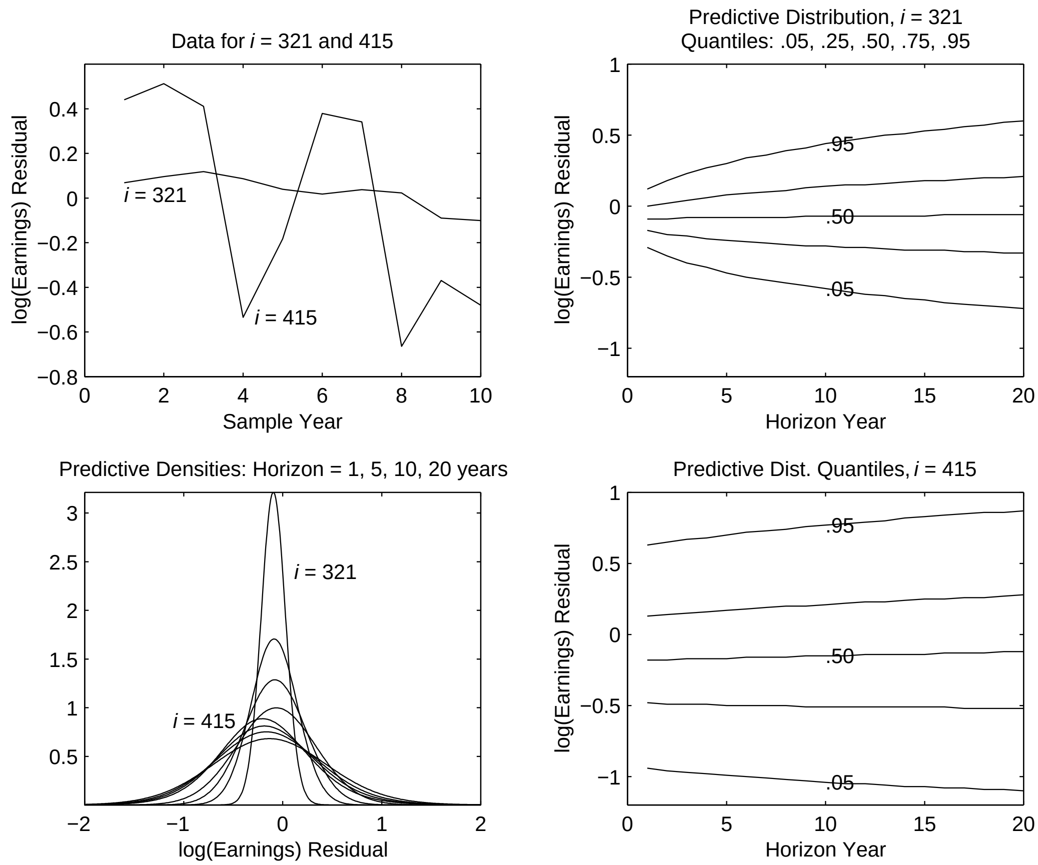

FIGURE 1.4. Data and predictive distributions for individuals $i = 321$ and 415.

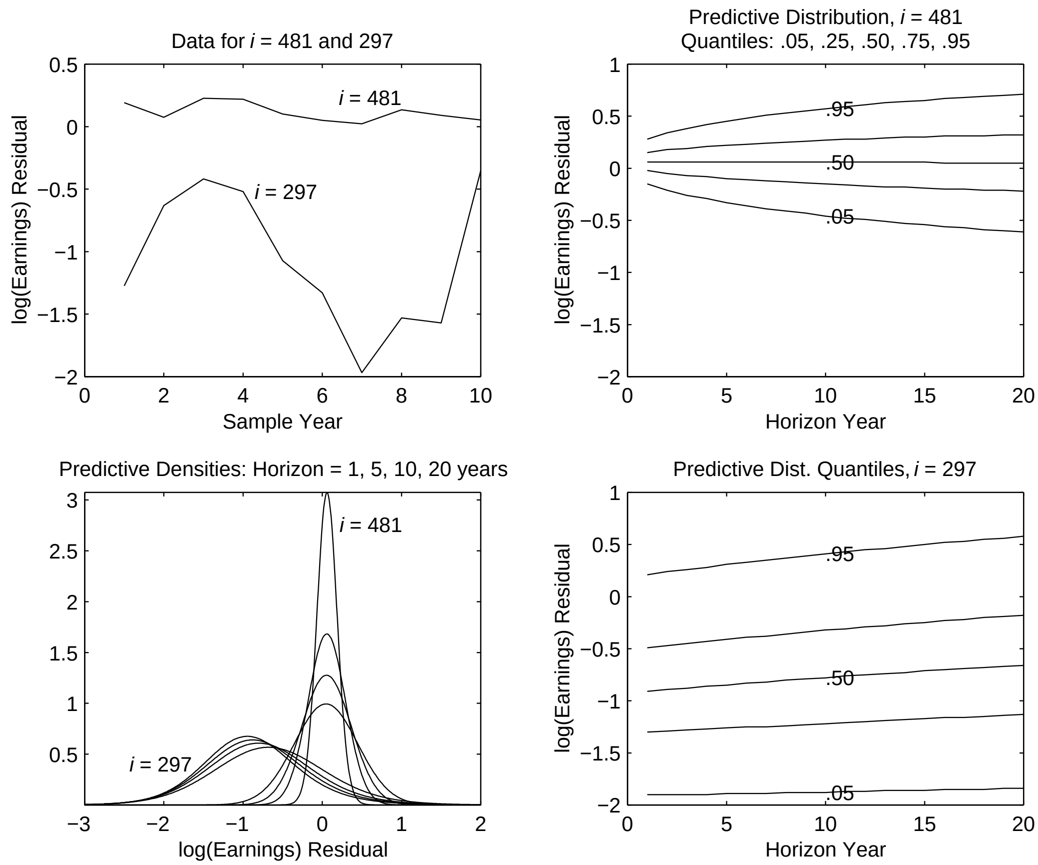

FIGURE 1.5. Data and predictive distributions for individuals $i = 481$ and 297.

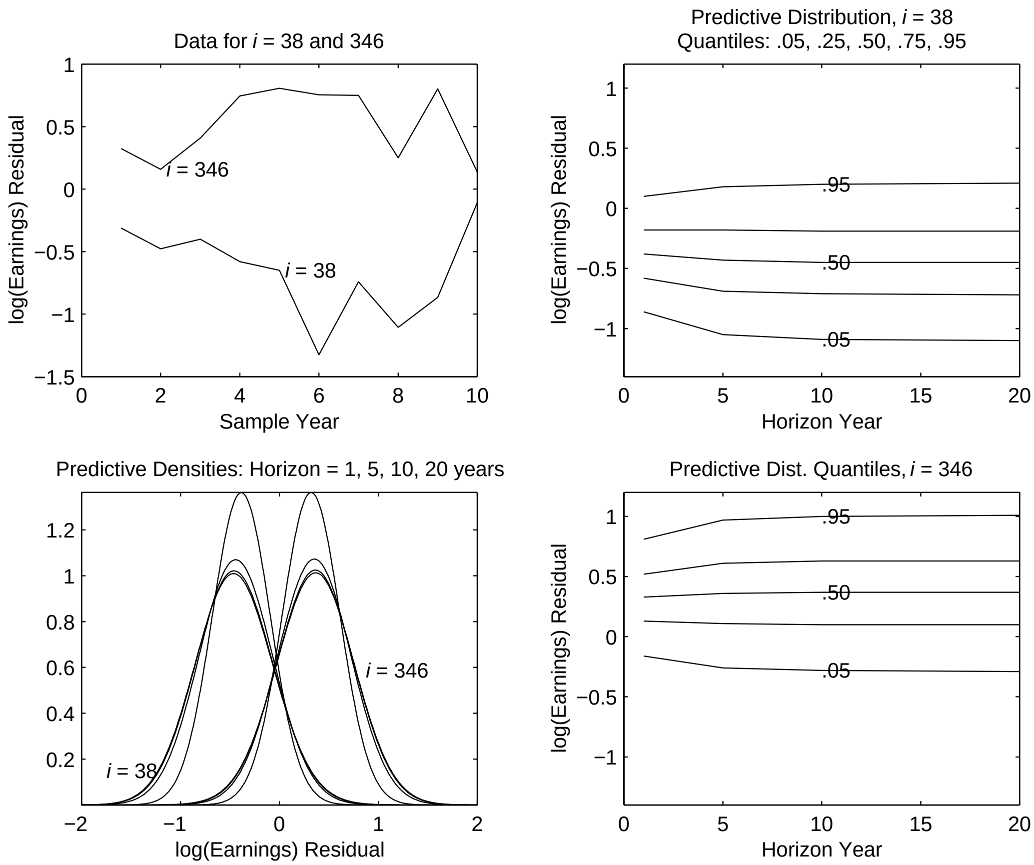

FIGURE 1.6. Data and predictive distributions for individuals $i = 38$ and 346; normal model with no heterogeneity in volatility (h_i constant).

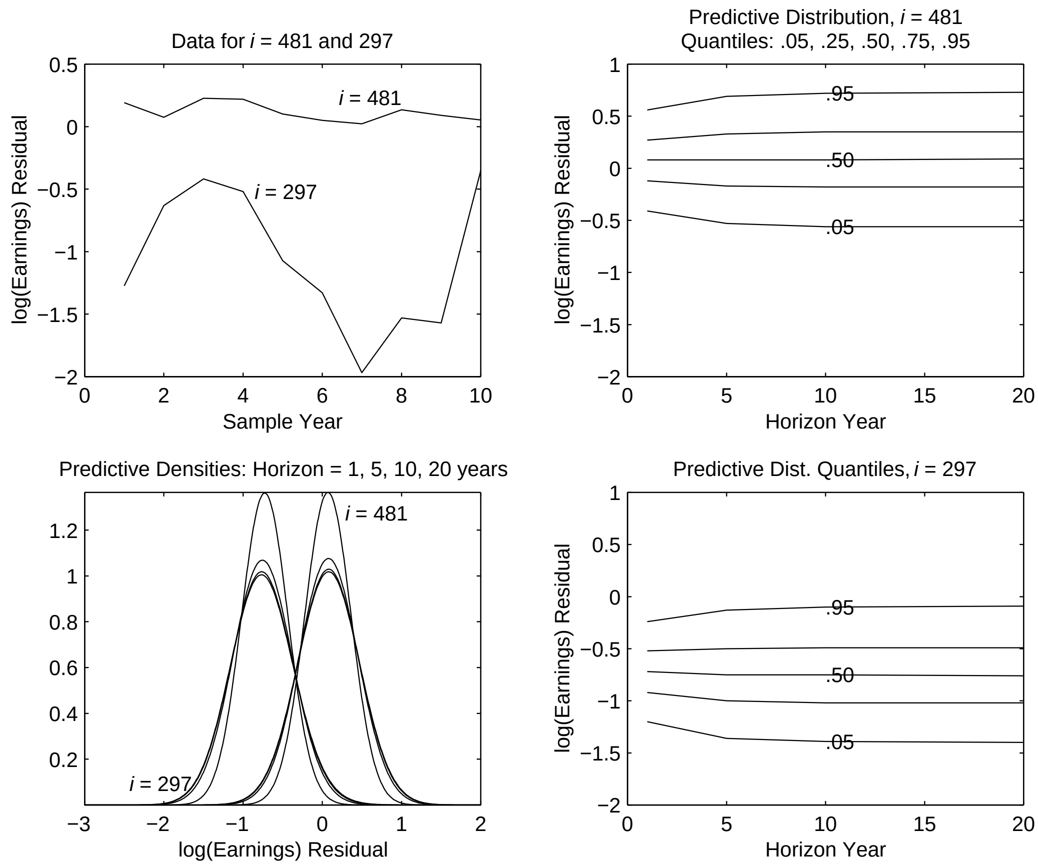

FIGURE 1.7. Data and predictive distributions for individuals $i = 481$ and 297; normal model with no heterogeneity in volatility (h_i constant).

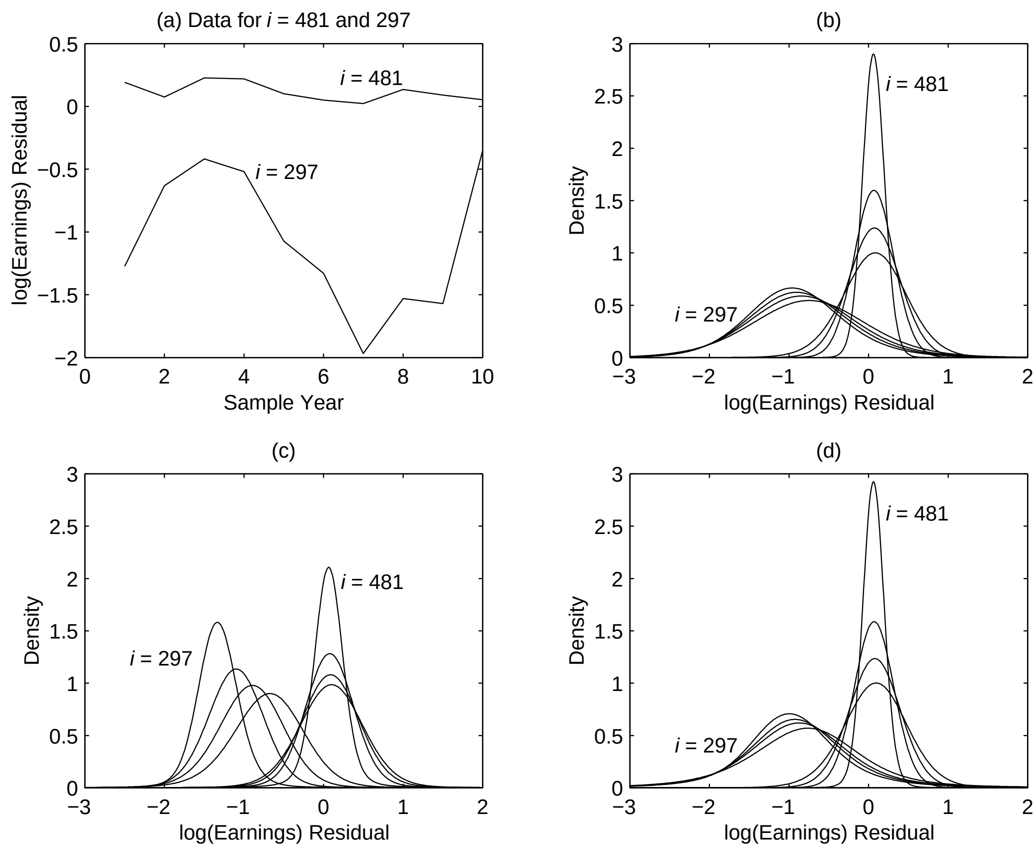

FIGURE 1.8. Predictive densities for Model (M_1), high-school sample: (b) $h_{it} = \phi_i$; (c) $h_{it} = \zeta_{it}/\tau$; (d) $h_{it} = \phi_i \zeta_{it}$.

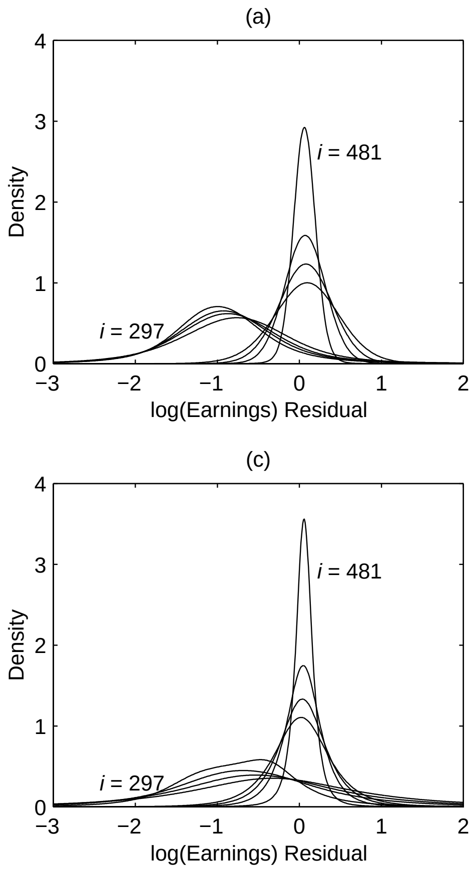

FIGURE 1.9. Predictive densities, high-school sample. (a) Model (M_1): $h_{it} = \phi_i \zeta_{it}$ (including α_i); (b) Model (M_1): $h_{it} = \phi_i \zeta_{it}$ (dropping α_i); (c) Model (M_2): $h_{it} = \phi_i \zeta_{it}$, $h_{wit} = \phi_{wi} \zeta_{wit}$; (d) Model (M_3): $h_{it} = \phi_i \zeta_{it}$, γ_i.

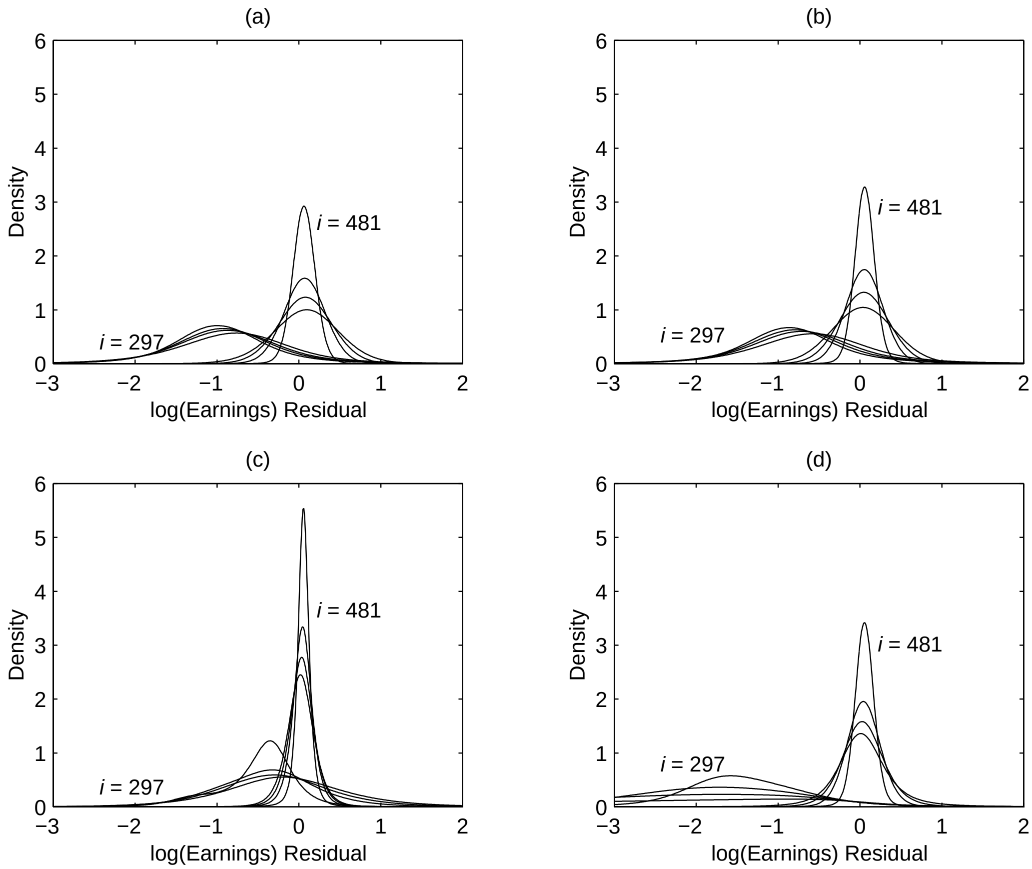

FIGURE 1.10. Predictive densities, college sample. (a) Model (M_1): $h_{it} = \phi_i \zeta_{it}$ (including α_i); (b) Model (M_1): $h_{it} = \phi_i \zeta_{it}$ (dropping α_i); (c) Model (M_2): $h_{it} = \phi_i \zeta_{it}$, $h_{wit} = \phi_{wi} \zeta_{wit}$; (d) Model ($M_3$): $h_{it} = \phi_i \zeta_{it}$, γ_i.