Penn Institute for Economic Research
Department of Economics
University of Pennsylvania
3718 Locust Walk
Philadelphia, PA 19104-6297
pier@econ.upenn.edu
http://www.econ.upenn.edu/pier

PIER Working Paper 01-023

"Learning-by-Doing as a Propagation Mechanism"

by

Yongsung Chang, Joao F. Gomes and Frank Schorfheide

Learning-by-Doing as a Propagation Mechanism

Yongsung Chang*, Joao Gomes†, Frank Schorfheide*‡

University of Pennsylvania

This Version: June 30, 2001

Abstract

This paper suggests that skill accumulation through past work experience, or "learning-by-doing", can provide an important propagation mechanism for initial shocks, as the current labor supply affects future productivity. Our econometric analysis uses a Bayesian approach to combine micro-level panel data with aggregate time series. Formal model evaluation shows that the introduction of the LBD mechanism improves the model's ability to fit the dynamics of aggregate output and hours.

JEL CLASSIFICATION: C52, E32, J22

KEY WORDS: Business-Cycle Fluctuations, Econometric Model Evaluation, Learning-by-Doing, Propagation of Shocks

---

*Department of Economics, 3718 Locust Walk, Philadelphia, PA 19104-6297.

†CEPR and Wharton School of Business, 3620 Locust Walk, Philadelphia, PA 19104-6302.

‡We thank Boyan Jovanovic, Donghoon Lee, James Nason, Richard Rogerson, participants at several workshops and conferences, Valerie Ramey (Co-Editor) and two anonymous referees for helpful comments and suggestions. The GAUSS programs that were used for the computations reported in the paper are available at *http://www.econ.upenn.edu/˜schorf*. An earlier version of this paper circulated under the title "Persistence."

1 Introduction

A well-known shortcoming of the standard dynamic general-equilibrium model, the so-called RBC model, is its weak internal-propagation mechanism. Aggregate output essentially traces out the movements of the exogenous technology process. This deficiency has been pointed out by Cogley and Nason (1995), among others. While GDP has an important trend-reverting component that is characterized by a hump-shaped response to a transitory shock (e.g., Blanchard and Quah (1989) and Cochrane (1994)), the standard RBC model invariably generates a monotonic response of output to transitory shocks. Furthermore, while output growth is positively autocorrelated in the data, the standard model cannot generate any persistence in output growth from a random-walk productivity process.

In this paper we suggest that allowing for a simple skill accumulation through past work experience, or "learning-by-doing," can help overcome some of these deficiencies. We will refer to the proposed model as the LBD model. Our point of departure from the standard model is motivated by a strong tradition in labor economics. Studies by Altug and Miller (1998) and Cossa, Heckman, and Lochner (2000) find a significant effect of past work experience on current wage earnings. It is also well documented that displaced workers suffer important wage losses (Jacobson, LaLonde, and Sullivan (1993)) and that wage profiles are affected by job tenure (Topel (1991)). These findings from the microeconomic data suggest that the aggregate economy experiences systematic changes in labor productivity, as business cycles are associated with strongly procyclical hiring of new workers and a countercyclical pattern of layoffs.

An important aspect of this paper is our methodology to estimate and evaluate the LBD model. To estimate the parameters of the skill accumulation process we derive a micro-level wage equation, whose structure closely resembles the aggregate wage equation. Rather than just using a simple "plug-in" technique, the micro estimates serve as a prior distribution for a Bayesian time series analysis of the aggregate LBD model. By scaling the prior covariance matrix, we can then control the relative weight of the microeconometric information. This procedure thus avoids the important criticism by Hansen and Heckman (1996) of the usage of incompatible micro estimates in highly aggregated macro models.

The evaluation of the aggregate LBD model is based on the framework proposed by Schorfheide (2000). It enables a formal comparison of population-moment and impulse-response-function predictions from dynamic stochastic general-equilibrium (DSGE) models to posterior estimates from vector autoregressions (VAR).[1] While both the log-linearized model and the VAR provide linear moving-average representations for aggregate data, the VAR representation is far less restrictive and is therefore suitable for a benchmark. Unlike in many previous studies, the benchmark for our impulse response function (IRF) comparisons is high. We are not simply comparing IRF shapes qualitatively. Instead, we examine whether the responses match quantitatively.

The main findings can be summarized as follows. First, introducing the LBD propagation mechanism improves the overall likelihood-based fit of the model relative to the standard RBC model. Second, the LBD model is able to generate a positive correlation in output growth, albeit a smaller one than in the data, even when the exogenous technology follows a random walk. Finally, the impulse-response function of output to a serially correlated transitory shock exhibits a pronounced hump shape, as the current increase in hours leads to a subsequent increase in labor productivity. While the LBD model is able to reproduce some important dynamics captured by the VAR reasonably well, it does retain some weaknesses. According to the model the response of hours is immediate, whereas it is delayed by 2–3 quarters in the data. Moreover, the model requires serially correlated external shocks to be able to generate pronounced hump-shaped responses.

A number of other studies have investigated the role of learning in generating richer macroeconomic dynamics. The study closest to this paper is Cooper and

[1] A survey of DSGE model evaluation procedures can be found in Kim and Pagan (1997). Alternative Bayesian approaches include DeJong, Ingram, and Whiteman (1996) and Geweke (1999a). A detailed comparison is provided in Schorfheide (2000).

Johri (1998). They included organizational capital in the production function and assumed that the current stock of organizational capital depends on past production rates. While this has obvious similarities with our learning mechanism, we believe that there are at least two advantages to our approach. First, by introducing learning through the direct effects of past work experience on current labor productivity we can measure its effects directly from panel data on wages and hours. Second, our modeling strategy avoids the thorny issue of distinguishing between internal and external learning-by-doing, and thus determining the components of national income to match up with the contribution of LBD. In our approach the benefits to learning are incorporated in workers' wages and thus are included in the labor share of national income.

Perli and Sakellaris (1998) and DeJong and Ingram (1998) emphasize schooling as a source of learning. Unlike our learning-by-doing, skill accumulation through schooling is costly due to fees and forgone market wages. While the level of schooling has a profound impact on long-run economic performance, most education takes place in the first 20 to 25 years of life. We believe that the wage-tenure link has been overlooked in previous studies and generates an at least equally compelling propagation mechanism for fluctuations at business cycle frequencies. Horvath (1999) proposes a role for learning in generating persistent effects from short-lived shocks. However, he relies on plant-level learning about productivities, which is less tractable and much more difficult to incorporate into DSGE models than our aggregate learning-by-doing mechanism.

Alternative propagation mechanisms have also been explored in various forms of adjustment costs in the allocation of labor, as in Cogley and Nason (1995), Burnside and Eichenbaum (1996), Andolfatto (1996), den Haan, Ramey and Watson (2000), Pries (1999), and Hall (1999).

The paper is organized as follows. Section 2 describes the proposed LBD model. Section 3 provides an overview of the econometric methodology employed in the paper. Section 4 is devoted to the estimation of the learning-by-doing effect using micro data obtained from the Panel Study of Income Dynamics (PSID). Section

5 contains the time-series analysis of the LBD model and our main findings. We conclude with a summary of our results and a discussion of some avenues for future research.

2 A Stochastic Growth Model with Learning-by-Doing

The proposed model economy is a variation of the standard stochastic-growth model. Our main departure is the introduction of a learning-by-doing mechanism associated with labor effort. The skill level of workers fluctuates over time according to their recent labor-supply history, which leads, in turn, to additional changes in output.

2.1 Households

The household sector is subsumed by a representative agent that maximizes expected discounted lifetime utility defined over consumption C and hours of work H:

$$U = \mathbb{E}_0 \left[\sum_{t=0}^{\infty} \beta^t [\ln C_t - B_t \frac{H_t^{1+1/\nu}}{1+1/\nu}] \right]. \tag{1}$$

Here $\mathbb{E}_0[.]$ denotes the expectation conditional on the information at time 0, while β is the subjective discount factor and B_t represents a stochastic shift of preferences. The assumption about the form of the momentary utility function is popular both in business-cycle analysis and the empirical labor-supply literature. The parameter ν denotes the compensated elasticity of labor supply. Log utility in consumption supports a balanced growth path.

Let X_t denote the experience of the representative agent from past labor supplies, which we identify with the skill level. We assume that the skill raises the effective unit of labor supplied by the household. According to our production technology, which will be specified below, labor input enters simply as an effective unit. Thus, a worker with skill X_t will earn a competitive wage rate of

$$W_t(X_t) = W_t^* X_t, \tag{2}$$

where W_t^* denotes the market wage rate for the efficiency unit of labor. The skill accumulates over time according to hours worked in the past. This dependence is summarized by the following representation:

$$\ln(X_t/X) = \phi \ln(X_{t-1}/X) + \mu \ln(H_{t-1}/H), \quad 0 \leq \phi < 1,\ \mu \geq 0, \tag{3}$$

where variables without time subscript denote steady states.

The household owns the capital stock K_t, and rents it to the firms at the competitive rental rate R_t. The budget constraint faced by the household is

$$C_t + I_t = W_t(X_t)H_t + R_t K_t, \tag{4}$$

where I_t denotes investment. Finally, the accumulation of capital is described by the law of motion:

$$K_{t+1} = I_t + (1-\delta)K_t, \tag{5}$$

where δ is the depreciation rate of the capital stock.

2.2 Firms

Firms produce final goods according to a constant-returns Cobb-Douglas technology in capital, K_t, and labor, N_t:

$$Y_t = K_t^{1-\alpha}(N_t A_t)^{\alpha}. \tag{6}$$

A_t denotes the exogenous stochastic technological progress. Reflecting the role of learning-by-doing, the labor input in production, N_t, consists of hours worked H_t, and the skill level X_t:

$$N_t = X_t H_t. \tag{7}$$

Profit maximization of the firm implies:

$$\begin{aligned} W_t^* &= \alpha A_t^{\alpha} N_t^{\alpha-1} K_t^{1-\alpha}, & (8) \\ R_t &= (1-\alpha)(N_t A_t)^{\alpha} K_t^{-\alpha}. & (9) \end{aligned}$$

2.3 Shocks

The model economy is pertubated by two exogenous processes: the level of technology A_t, and the preference process B_t. We assume that the production technology evolves according to a random walk with drift:

$$\ln A_t = \gamma + \ln A_{t-1} + \epsilon_t^a, \quad \epsilon_t^a \sim \mathcal{N}(0, \sigma_a^2). \tag{10}$$

Thus, the shock ϵ_t^a has a permanent effect on the level of productivity.

The preference shock B_t shifts the marginal rate of substitution between goods and leisure, which has been emphasized by Parkin (1988), Baxter and King (1991), and Hall (1997) as an important factor in explaining aggregate labor-market fluctuations. We assume that B_t follows a stationary AR(1) process:

$$\ln B_t = (1-\rho)\ln B + \rho \ln B_{t-1} + \epsilon_t^b, \quad \epsilon_t^b \sim \mathcal{N}(0, \sigma_b^2), \tag{11}$$

where $0 \leq \rho < 1$. The preference shock ϵ_t^b has only a transitory effect. While preference shocks are obviously quite convenient and parsimonious, they are not crucial for our findings. The results are robust to allow for alternative types of transitory shocks such as government spending shocks and temporary movements in productivity.[2]

The two shocks ϵ_t^a and ϵ_t^b are assumed to be uncorrelated. Suppose a bivariate VAR with sufficiently many lags is fitted to data on output growth and hours generated from the LBD model. The presence of one permanent and one transitory shock allows us to use the Blanchard and Quah (1989) scheme to identify the two structural shocks.

2.4 Discussion

Several studies in labor economics have documented the role of past labor supply in determining current wages. Altug and Miller (1998) in particular find a significant

[2]With essentially the same parameter values, we are also able to generate a pronounced hump-shaped response in output in response to transitory shifts in government spending or productivity.

learning-by-doing effect of past labor supplies on current wages. There is also ample evidence of a strong job-tenure effect in wage profiles (Topel (1991)) and significant wage losses suffered by displaced workers (e.g., Jacobson, LaLonde, and Sullivan (1993)).

Although we will continue to refer to these results, and others, as evidence of learning-by-doing, perhaps more appropriately we should regard them merely as evidence for past labor supply in determining current wages. Whether this suggests an important role for learning-by-doing or simply reflects, for example, a loss of employment rents by displaced workers as emphasized in the matching literature, is not crucial to our purposes. The key implication of our model is only that there is a strong link between past labor supply and current wages.

Our parsimonious representation of skill accumulation in Equation (3) provides a simple interpretation of the learning-by-doing mechanism. First, it implies that an increase in the number of hours worked in the current period contributes to an improvement in labor skills in the next period with an elasticity of μ. Second, skill accumulation is persistent but not permanent. The effects on skills decay over time at rate ϕ.

Finally, note that the LBD model is identical to the standard stochastic-growth model if $\mu = 0$ and $X_t = X$. Hence, it is quite easy to understand the different implications of the two models, in particular with respect to their dynamic responses to exogenous shocks. While the RBC model fails to generate much internal propagation, our LBD model has the potential to generate richer dynamics. Workers' increased effort in expansions leads to a rise in future productivity through the accumulation of skills, and to corresponding increases in output.

3 Quantitative Analysis: Methodology

The properties of the LBD model crucially depend on the parameterization of the skill-accumulation process. Data from the PSID are used to estimate the effect of learning-by-doing on the micro level. Taking these micro estimates into account, the representative-agent model is fitted to quarterly time-series data on GDP growth and aggregate hours – two important aggregate quantities. Moreover, the use of output and hours makes our work comparable to earlier studies such as Blanchard and Quah (1989) and Cogley and Nason (1995). In addition to the time-series fit of the model, we examine how well the LBD specification captures the business-cycle dynamics of output growth and hours based on impulse-response and autocorrelation functions.

3.1 Combining Micro and Macro Data

A Bayesian approach, placing probability both on data and on models, is adopted for our econometric analysis. The use of prior distributions enables us to incorporate external information into the parameter estimation. Two data sets are used in the empirical analysis: micro-level PSID data on wages and hours (denoted by Y^m) and macro-level time-series Y^M on output growth and aggregate hours.

The LBD model provides a joint-probability distribution $p(Y^M|\theta^{LBD}, \theta^M)$ for aggregate time-series conditional on the learning-by-doing parameters $\theta^{LBD} = [\mu, \phi]'$ and additional parameters θ^M that appear in the representative-agent model. A log-wage equation for individual i,

$$w_{i,\tau} = \tilde{w}^*_\tau + x_{i,\tau} + q_{i,\tau}, \tag{12}$$

is used to obtain a probabilistic representation for the micro data $p(Y^m|\theta^{LBD}, \theta^m)$. The terms w^*_τ and $x_{i,\tau}$ are micro counterparts of $\ln W^*_t$ and $\ln X_t$, respectively. Individual-specific components of the wage are captured by $q_{i,\tau}$. Unlike the aggregate time-series data, which are available at quarterly frequency, the panel data are observed only at annual frequency. We use τ instead of t as an annual time index. As in the representative-agent model, the skill stock $x_{i,\tau}$ of individual i is a moving average of hours worked in the past:

$$x_{i,\tau} = \tilde{\mu} \sum_{j=0}^{\infty} \tilde{\phi}^j h_{i,\tau-1-j}, \tag{13}$$

where h is log annual hours. Since $x_{i,\tau}$ is an annual process, we denote the parameters of its law of motion by $\tilde{\mu}$ and $\tilde{\phi}$. The Appendix describes how the annual parameters can be converted into quarterly μ and ϕ. Details including the stochastic process $q_{i,\tau}$ will be discussed in Section 4.

In principle, the micro data Y^m contain some information on the aggregate price W_t^* for an efficiency unit of labor. However, in our subsequent analysis we neglect this link and assume

$$p(Y^M|Y^m, \theta^{LBD}, \theta^M) \approx p(Y^M|\theta^{LBD}, \theta^M). \tag{14}$$

Since the micro data are annual, they convey much less information on the business-cycle variation of the aggregate skill price than do the quarterly time-series observations. We believe that potential gains from using the exact conditional distribution are too small to justify the additional computational burden.

Under the above assumption the joint posterior distribution for θ^{LBD} and θ^M simplifies to

$$p(\theta^{LBD}, \theta^M|Y^M, Y^m) = \frac{p(Y^M|\theta^{LBD}, \theta^M)p(\theta^M|\theta^{LBD})p(\theta^{LBD}|Y^m)}{\int p(Y^M|\theta^{LBD}, \theta^M)p(\theta^M|\theta^{LBD})p(\theta^{LBD}|Y^m)dY^A}, \tag{15}$$

where $p(\theta^M|\theta^{LBD})$ is the prior distribution for the additional parameters of the LBD model conditional on μ and ϕ. Equation (15) implies that the empirical analysis can proceed in two steps: (i) estimate the LBD parameters based on the panel data Y^m to obtain the *posterior* $p(\theta^{LBD}|Y^m)$; and (ii) use the micro estimate $p(\theta^{LBD}|Y^m)$ as a *prior* distribution for the LBD parameters in the time-series estimation of the aggregate model.

3.2 Evaluation of Aggregate Models

Three structural aggregate models are considered. First, we report evaluation statistics for the RBC model ($\mu = 0$, $\phi = 0$) as a basis for comparisons. Second, the aggregate LBD model is combined with the prior distribution for μ and ϕ from the panel-data analysis. Third, another version of the LBD specification is obtained by scaling the prior variance for μ and ϕ to assign less weight to the micro estimates.

This provides an interesting opportunity to examine a possible tension between the micro and macro estimates of θ^{LBD} because under the diffuse prior the learning-by-doing parameters can adjust to match the dynamics of the LBD model with the aggregate time-series data. (Details are discussed in Section 5.) We will refer to the three structural specifications as DSGE models. The models are log-linearized and solved by standard methods, such that of King, Plosser, and Rebelo (1988). Since the models are driven by two exogenous processes, A_t, and B_t, the marginal distribution of output growth and hours is non-degenerate.

The parsimonious and stylized nature of the model economies is a potential source for misspecification. To account for this problem a vector autoregression serves as a reference model. VARs are widely used in empirical macroeconomics to study the dynamic properties of multivariate time series. While both the DSGE models and the VAR imply a linear moving-average representation for aggregate output growth and hours, the VAR is far less restrictive.

The posterior odds of the DSGE models versus the VAR provide a measure of the overall statistical fit of the DSGE models within the class of bivariate linear time-series models. If the statistical fit of the DSGE models is poor, then the VAR can be used as a benchmark to obtain posterior estimates of population moments and impulse response functions. The model evaluation can proceed by comparing the predictions of the DSGE models and the posterior VAR estimates with respect to population characteristics of interest, denoted by an $m \times 1$ vector φ. A formal interpretation of this procedure and a detailed comparison with other evaluation approaches such as calibration and likelihood-based model comparisons can be found in Schorfheide (2000). To make the paper self-contained, a brief description of the evaluation steps is provided.

Denote the reference model by $\mathcal{M}_0$ and the structural models by $\mathcal{M}_i$, $i = 1, \dots, k$ with parameters $\theta_{(i)}$. Let $\pi_{i,0}$ be the prior probability of model $\mathcal{M}_i$ (conditional on the micro data Y^m) and $\pi_{i,T}$ its posterior probability. The population characteristics φ are functions $f_i(\theta_{(i)})$ of the model parameters $\theta_{(i)}$. Their posterior conditional on a model $\mathcal{M}_i$ is denoted by the density $p(\varphi|Y, \mathcal{M}_i)$, where Y denotes the combination of the micro data Y^m and aggregate data Y^M.

Step 1: Compute posterior distributions for the model parameters $\theta^{(i)}$ and calculate posterior model probabilities:

$$\pi_{i,T} = \frac{\pi_{i,0} p(Y^M|Y^m, \mathcal{M}_i)}{\sum_{i=0}^{k} \pi_{i,0} p(Y^M|Y^m, \mathcal{M}_i)}, \tag{16}$$

where $p(Y^M|Y^m, \mathcal{M}_i) = \int p(Y^M|\theta_{(i)}, \mathcal{M}_i) p(\theta_{(i)}|Y^m) d\theta_{(i)}$ is the marginal data density of $\mathcal{M}_i$. The posterior probabilities provide a measure of time-series fit for the models $\mathcal{M}_i$.

Step 2: The overall posterior distribution of the population characteristics φ is given by

$$p(\varphi|Y) = \sum_{i=0}^{k} \pi_{i,T} p(\varphi|Y, \mathcal{M}_i). \tag{17}$$

If the posterior probability of the reference model is substantially larger than the posterior probabilities of the structural models, that is, $\pi_{0,T} \gg \pi_{i,T}$, $i = 1, \ldots, k$, then

$$p(\varphi|Y) \approx p(\varphi|Y, \mathcal{M}_0). \tag{18}$$

Step 3: Loss functions $L(\varphi, \hat{\varphi}_{i,b})$ are introduced to penalize the deviation of actual model predictions $\hat{\varphi}_{i,b}$ (based on structural Bayes estimates) from population characteristics φ. For each structural model $\mathcal{M}_i$, we examine the expected loss associated with $\hat{\varphi}_{i,b}$ under the posterior distribution of φ conditional on the VAR:

$$R(\hat{\varphi}_{i,b}|Y, \mathcal{M}_0) = \int L(\varphi, \hat{\varphi}_{i,b}) p(\varphi|Y, \mathcal{M}_0) d\varphi. \tag{19}$$

The loss function can be used in several ways to analyze the model. First, the expected loss $R(\hat{\varphi}_{i,b}|Y, \mathcal{M}_0)$ itself provides an absolute measure of fit. Second, the differential across structural models provides a relative measure of fit that allows model comparisons.

Two loss functions are used in the empirical analysis. The quadratic loss function is of the form

$$L_q(\varphi, \hat{\varphi}) = (\varphi - \hat{\varphi})' W (\varphi - \hat{\varphi}), \tag{20}$$

where W is a positive definite $m \times m$ weight matrix. It can be shown that under the quadratic loss function the ranking of model predictions $\hat{\varphi}$ depends only on the weighted distance

$$\tilde{R}_q(\hat{\varphi}|Y) = (\hat{\varphi} - I\!E[\varphi|Y])' W (\hat{\varphi} - I\!E[\varphi|Y]) \tag{21}$$

between predictor $\hat{\varphi}$ and the overall posterior mean $I\!E[\varphi|Y]$.

The second loss function penalizes predictions that lie far in the tails of the overall posterior distribution $p(\varphi|Y)$. Define

$$C_{\chi^2}(\varphi|Y) = (\varphi - I\!E[\varphi|Y])' Var^{-1}[\varphi|Y](\varphi - I\!E[\varphi|Y]) \tag{22}$$

and let

$$L_{\chi^2}(\varphi, \hat{\varphi}) = \mathcal{I}\Big\{ C_{\chi^2}(\varphi) < C_{\chi^2}(\hat{\varphi}) \Big\}, \tag{23}$$

where $I\{x < x'\}$ is equal to one if $x < x'$, and is equal to zero otherwise. The expected L_{χ^2} loss is similar to a p-value if the posterior density is well approximated by a unimodal Gaussian density. A value close to one indicates that the DSGE model predictions lie far in the tails of the overall posterior distribution.

A comparison of the DSGE models based on posterior odds ratios obtained in *Step 1* can be rationalized through a loss function that assigns the loss 1 if the "false" model is chosen, and 0 if the "correct" model is chosen. However, unlike the previous two loss functions, the 0–1 loss function does not capture the economic implications of selecting one of the structural models, if the reference model $\mathcal{M}_0$ is "correct."

4 Micro Evidence on Skill Accumulation

Panel data from the PSID for the period 1971-1992 are used to estimate the wage equation

$$w_{i,\tau} = \tilde{w}^*_\tau + \tilde{\mu} \sum_{j=0}^{\infty} \tilde{\phi}^j h_{i,\tau-1-j} + q_{i,\tau}. \tag{24}$$

The ideosyncratic productivity $q_{i,\tau}$ is decomposed as follows

$$q_{i,\tau} = \zeta_i + v'_{i,\tau}\lambda_v + u_{i,\tau}, \tag{25}$$

where $v_{i,\tau}$ is a 4×1 vector of covariates that helps explain wage differentials across individuals. It consists of age and squared age to capture the age-earnings profile (AEP), years of schooling, and a female dummy. ζ_i represents an individual-specific effect, and $u_{i,\tau}$ generates stochastic shifts of the idiosyncratic productivity. We allow for autocorrelation in $u_{i,\tau}$

$$u_{i,\tau} = \xi u_{i,\tau-1} + \epsilon^u_{i,\tau}, \quad \epsilon^u_{i,\tau} \sim iid\mathcal{N}(0, \sigma^2_u). \tag{26}$$

For the actual estimation we truncate the infinite-order moving average of past wages after 5 lags and replace the market wage rate for efficiency unit of labor $\tilde{w}^*_\tau$ by a $(\Upsilon + 1) \times 1$ vector of time dummies d^0_τ. This leads to

$$w_{i,\tau} = \zeta_i + d^{0'}_\tau \gamma_d + v'_{i,\tau}\gamma_v + \tilde{\mu} \sum_{j=0}^{4} \tilde{\phi}^j h_{i,\tau-1-j} + u_{i,\tau}, \quad \tau = 0, 1, \ldots, \Upsilon. \tag{27}$$

To construct a Gaussian likelihood function for wages conditional on hours[3] and covariates $v_{i,\tau}$, Equation (27) is quasi-differenced[4]:

$$\begin{aligned} w_{i,\tau} \;&=\; \xi w_{i,\tau-1} + \tilde{\zeta}_i + d'_\tau \tilde{\lambda}_d + v'_{i,\tau}\tilde{\lambda}_v \qquad (28) \\ &\quad + \tilde{\mu} h_{t-1} + \sum_{j=1}^{4} \tilde{\mu}(\tilde{\phi}^j - \xi \tilde{\phi}^{j+1}) h_{t-1-j} - \tilde{\mu}\xi\tilde{\phi}^5 h_{t-6} + \epsilon^u_{i,\tau}, \quad \tau = 1, \ldots, \Upsilon. \end{aligned}$$

Subsequently, we will briefly report our prior distribution for the coefficients that appear in the wage equation, discuss the parameter estimates, and compare our estimates to the existing literature.

[3]Even if the hours innovations and wage innovations are correlated, the conditional-likelihood function generates the same estimates for the parameters of the wage equation as the joint estimation of a VAR for hours and wages if: (i) the conditional mean of log hours is modeled as a linear function of the regressors that appear in the wage equation; and (ii) the coefficients in the hours equation are integrated out with respect to a prior that is uniform on the real line.

[4]Unlike d^0_τ, the vector d_τ does not include a dummy variable for $\tau = 0$. The transformed parameters satisfy the relationship $\tilde{\zeta}_i + v'_{i,\tau}\tilde{\lambda}_t = \zeta_i(1 - \xi) + (v'_{i,\tau} - \xi v'_{i,\tau-1})\lambda_v$.

4.1 Priors for Micro-Level Estimation

An uninformative prior is used for the LBD parameters $p(\tilde{\mu}, \tilde{\phi}) \propto 1$. This implies that we simply explore the shape of the marginal likelihood function in the direction of μ and ϕ. The prior for the autoregressive coefficient ξ and the standard deviation σ_u of the random effect is of the form $p(\xi, \sigma_u) \propto \sigma_u^{-1}$.

Our prior distributions for the parameters ζ_i, λ_d, and λ_v are provided in Table 1. The priors are centered around estimates that have been reported in the empirical micro literature based on data sets other than the PSID. *A priori* the average "return-to-schooling" is 10% (e.g., Willis (1986)) the gender bias is -40% (e.g., Cain (1986)). The age-earnings profile peaks around at the age of 55 and the individual at his peak earns 40% more than he did at the age of 30 (e.g., Ghez and Becker (1975)). We use fairly large standard deviations so that our priors assign substantial probabilities to ranges that most economists would regard as plausible, and permit some unreasonable values as well.

The prior for the individual-specific effect is $\mathcal{N}(0, 0.25^2)$. This implies that among individuals with the same age, education, and employment history the wage ratio of the bottom 2.5% and the top 97.5% quantile of the distribution is about 1/3. Finally, consider a baseline individual, who is a 30-year-old male in 1976 with 12 years of schooling. *A priori* his log real wage in 1983 dollars is $\mathcal{N}(\ln 10, 0.2^2)$. To generate a prior for the coefficients on the time-dummy variables, we assume that it subsequently evolves in steady state according to $\ln w_\tau = \gamma + \ln w_{\tau-1} + \epsilon_\tau^w$, where $\gamma \sim \mathcal{N}(0.02, 0.004^2)$ and $\epsilon_\tau^w \sim iid\mathcal{N}(0, 0.01^2)$. The prior on wage growth corresponds to the prior on productivity growth that is used for the time-series estimation.

4.2 Posterior Estimates from the PSID

Posterior means and standard errors for the LBD parameters and some of the additional parameters are reported in Table 1. Computational details are provided in the Appendix. According to our estimates the age-earnings profile peaks at age 59, the "return-to-schooling" is 8.7%, and women earn about 30% less than men. These estimates indicate that the inclusion of lagged hours in the wage equation does not lead to unreasonable estimates of other widely studied determinants of wages at the micro level cited above.

Of primary interest are the estimates of the learning-by-doing parameters, $\tilde{\mu}$ and $\tilde{\phi}$. Based on the annual panel data, the posterior means for the two LBD parameters are $\tilde{\mu} = 0.33$ and $\tilde{\phi} = 0.41$. Suppose individual i works on average 2000 hours per year and attained her steady-state skill level in 1980. Due to the recession in 1981 she becomes unemployed for three months and the annual hours drop to 1500 hours. In the subsequent years she will again be employed for 2000 hours. Our estimates imply that individual i's wage will drop by 9.5% in 1982 due to the loss of work experience. However, in 1983 her wages will be only 3.9% below the steady-state level, and by 1985 she has almost completely recovered from the temporary loss of experience as her wage is below the steady state by just 1.6%.

In order to use the LBD parameter estimates in the quarterly time-series analysis, $\tilde{\mu}$ and $\tilde{\phi}$ are converted into μ and ϕ according to the formula derived in the Appendix. The values at quarterly frequency are $\mu = 0.111$ and $\phi = 0.798$, with standard errors of 0.004 and 0.012.[5]

We compare our estimates to those of several other studies. Using PSID data for women Altug and Miller (1998) estimate that the wage elasticity of lagged hours is 0.2 and the wage elasticity of hours lagged twice is 0.05. In our notation their estimates correspond to $\tilde{\mu} = 0.2$ and $\tilde{\phi} = 0.25$, implying somewhat weaker initial impact and less persistence of its subsequent effects than our estimates. Based on the PSID as well, Topel (1991) finds evidence of a clear tenure effect that leads to a wage growth of about 7% after one year of tenure, leveling off to around 2.5% after 10 years of work.

[5] The estimates are fairly robust to reasonable changes in the prior distribution. Changes in the mean of the baseline wages to $\ln 8$ or $\ln 12$ had virtually no effect. Scaling the prior standard deviation of the individual-specific effect by 0.1 increased the estimate of μ to 0.13. An increase of the prior variance by a factor of 10 reduced the estimate of μ to 0.09. The estimates of ϕ remained between 0.80 and 0.83.

The effects of job separations on wages are also well documented although quantitative estimates are quite sensitive to the exact definition of separation. Very strict definitions of job separations usually lead to estimates of subsequent wage losses that are both high and persistent, while broader definitions usually yield much smaller numbers. For example, Jacobson, LaLonde, and Sullivan (1993) find wage losses between 10% and 25% in the first year following a separation. In our model, this implies an annual value of μ of between 0.1 and 0.25. The time for full recovery varies across sample. For the mass-layoff sample, wage loss is smaller and the recovery is fast, suggesting small values for both μ and ϕ. For non-mass layoffs, wage losses are much larger and the recovery is slow, suggesting higher values for both μ and ϕ.

We adopted a broad notion of learning-by-doing in our panel-data estimation because it is compatible with the representative-agent model. However, we generate a second prior for the time-series estimation of the LBD model by scaling the covariance matrix of the micro estimates. Confidence sets constructed from this diffuse prior will cover the alternative estimates that have been reported in the literature.

5 Time-Series Estimation and Aggregate Model Evaluation

A common approach in the calibration literature is to evaluate models based on parameter values that are regarded as economically plausible. Such values are obtained from long-run averages of aggregate time series, from microeconometric evidence, or from pure introspection. In our analysis the micro-level information is formally incorporated through a prior distribution. Moreover, we will specify priors for the remaining structural parameters, which can be justified based on a short sample of observations preceding our estimation period. The priors are combined with likelihood functions for the DSGE models to obtain posterior distributions. Loosely speaking, the Bayes estimation can be interpreted as searching for parameter values such that the DSGE models fit the data in a likelihood sense, without deviating too much from economically sensible values.

All models are fitted to quarterly U.S. data from 1954:III to 1997:IV. Priors are specified conditional on the first 22 observations. The estimation period is then 1960:I to 1997:IV. A fourth-order VAR with "Minnesota Prior" (Doan, Litterman, and Sims (1984)) serves as the reference model. Data definitions and some computational details are provided in the Appendix.

5.1 Prior Distributions for the DSGE Model Parameters

The joint posterior distribution for LBD parameters μ and ϕ obtained from the panel-data analysis in Section 4 is approximated by a bivariate normal distribution, which serves as Prior 1 for the LBD model. A second prior, denoted as Prior 2, is obtained by scaling[6] the covariance matrix of Prior 1 by 25^2. Under Prior 2, the mean of μ is 0.145 and the standard error increases to 0.078. The mean of ϕ is 0.664 and its standard error is 0.213. The second prior implicitly assigns less weight to the micro-level evidence on learning-by-doing. Effectively, this provides us an opportunity to investigate a possible tension between the learning mechanism at the micro level and the propagation mechanism needed to fit the time-series data at the aggregate level. For the RBC model, we set $\mu = \phi = 0$.

The prior distributions for the remaining parameters are summarized in Table 2. The shapes of the densities are chosen to match the natural domain of the structural parameters. It is assumed that the parameters are *a priori* independent of μ, ϕ, and each other. Based on pre-sample information on the labor share of national income, the prior for α is centered at 0.66. The prior for β implies that the steady-state real interest rate is about 4% per year. The economy is expected to grow an average 2% annually. Capital is assumed to depreciate at the rate of 2.5% quarterly. As these parameters are standard in the literature, we use fairly small standard errors to make our model *a priori* comparable to those in the literature.

[6] We also truncate the distribution to guarantee that $\mu \geq 0$ and $0 \leq \phi < 1$.

The prior for the aggregate short-run labor supply elasticity υ is centered at 2 as in Prescott (1986), with a standard error of 0.5. We use diffuse priors for the parameters of the exogenous processes.[7] The 95% confidence interval for the autocorrelation ρ of the preference process B_t extends from 0.6 to 0.99. For the standard deviations σ_a and σ_b of the structural shocks, we use uninformative inverse-Gamma priors.

5.2 Posterior Estimates

Draws from the posterior distribution of the DSGE model parameters are generated with a random walk Metropolis algorithm. The posterior means and standard errors are reported in Table 3. As can be seen in the first column, parameter estimates for the RBC model are similar to those in the literature. The model requires a fairly persistent preference shock because its estimated autoregressive coefficient is 0.944. This reflects the high autocorrelation present in the aggregate-hours series.[8]

Of primary interest to our analysis are the learning-by-doing parameters. Confidence sets for μ and ϕ are plotted in Figure 1. The top graph shows the results for Prior 1. Prior and posterior confidence sets are almost identical as our priors are informative. The posterior means of μ and ϕ are 0.111 and 0.797, respectively. Prior 2 is much more diffuse than Prior 1 and the likelihood function pulls the posterior toward a higher value of μ and a lower value of ϕ. The posterior estimates $\hat{\mu} = 0.145$ and $\hat{\phi} = 0.664$ imply a somewhat stronger initial impact and faster recovery toward the steady state. The estimation results suggest some discrepancy between the information in the panel data and the parameter values that are needed to achieve the best time-series fit of the LBD model. We will further examine this discrepancy in the context of the impulse-response function analysis. The other

[7] To fit the models to the data, we parameterize them in terms of log steady-state hours h, rather than in terms of the mean preference level B. Since h does not affect the dynamics of the models, we do not report it in Tables 2 and 3.

[8] It is well known that aggregate hours series exhibit large persistence. For example, Shapiro and Watson (1988) emphasize permanent shifts in labor supply as an important source of economic fluctuations.

parameters are similar across all three models. The only exception is that the LBD models require slightly less persistent preference shock to fit the time series since the learning-by-doing provides an internal propagation mechanism.

Formal statistics of overall time-series fit – namely marginal data densities and posterior model probabilities for the three DSGE model specifications and the VAR – are summarized in Table 4. The marginal data densities $\ln p(Y^M|Y^m, M_i)$ can be interpreted as maximum log-likelihood values, penalized for model dimensionality, and adjusted for the effect of the prior distribution. The posterior odds ratios of the LBD specifications versus the RBC model indicate that the learning-by-doing propagation mechanism clearly improves the time-series fit of the stochastic growth model. The LBD models are favored at rates of 4 to 1 and 20 to 1, respectively. However, the results also indicate that the reference model outperforms the structural models by a wide margin. Suppose the prior probabilities are 0.25 for each of the four specifications. In this case the reference model has essentially posterior probability one. This result is robust across a wide range of prior probabilities. To shed more light on how well the DSGE models capture the dynamics of output growth and hours, we examine the implied impulse-response functions and autocorrelations in the next section.

5.3 Impulse-Response Dynamics

The dynamic behavior of DSGE models is often summarized by their impulse-response functions. One of the well-known time-series properties of output in a VAR analysis is its hump-shaped, trend-reverting response to a transitory shock. This has been documented by, among others, Blanchard and Quah (1989) and Cochrane (1994), and it has been one of the main criticisms of the simple RBC model. Cogley and Nason (1995) comment that "while GNP first rises and then falls in response to a transitory shock, the RBC model generates monotonic decay. Thus, the model does not generate an important trend-reverting component in output."

In our analysis, the model economies are driven by a random-walk technology and a stationary preference shift. The innovations in the technology process have a permanent effect on output, whereas the innovations in the preference process have a transitory effect. Blanchard and Quah's (1989) method is used to identify transitory and permanent shocks in the vector autoregression. We then assess the discrepancy between the different DSGE model impulse responses and the VAR response functions.

Figure 2 depicts posterior mean responses for the three DSGE model specifications and the VAR, as well as a pointwise 75% Bayesian highest posterior density (HPD) confidence band from the VAR. All responses are based on one-standard-deviation shocks. The first two rows are the responses of output and hours to a transitory shock, and the last two rows are the responses to a permanent shock.

The VAR responses of output to the transitory shock exhibit a clear hump shape. In the context of the DSGE models, the transitory shock causes an increase in the marginal rate of substitution between goods and leisure, that is, a decrease in B_t. The upper-left panel shows that after the impact of the shock both output and hours are monotonically decreasing in the RBC model. Columns 2 and 3 of Figure 2 depict the responses of the LBD economies with Prior 1 and Prior 2, respectively. Unlike in the standard model, output rises for the first four (Prior 2) to eight (Prior 1) quarters and decays subsequently. The decay in output is somewhat slower than the decay predicted by the VAR impulse-response function. Under Prior 2 the posterior estimate of ϕ is smaller, so the decay is slightly faster. This problem of slow decay is also apparent in the IRF plots generated by Perli and Sakellaris (1998), but is less severe for our IRFs. We will show subsequently that there exists a parameterization of the LBD model that overcomes this deficiency, although one has to deviate from the time-series estimates to achieve it.

While a hump-shaped output response is obtained for both LBD specifications, only the LBD model with Prior 2 is capable of generating a very small one-period hump in the response of hours. However, according to the VAR the hump in the hours dynamics is as pronounced as the one in the output IRF. In response to a permanent shock, the RBC model and the LBD model with Prior 1 deliver monotone IRFs, whereas the VAR indicates that both hours and output respond with a delay.

Again, only the LBD model with Prior 2 generates a very slight hump in the response of hours to a permanent technology shock.

5.4 The Effect of ρ, μ, and ϕ on the Propagation of Shocks

To illustrate the sensitivity of the impulse responses to changes in the values of the learning-by-doing parameters μ and ϕ and the autocorrelation of the transitory shock ρ, we explore alternative parameterizations of the LBD model.

Consider the effects of a transitory preference shock. Upon impact the labor supply rises. In Period 2 and thereafter two competing effects occur simultaneously. On the one hand, the preference process reverts to its long-run mean, which decreases the labor supply. On the other hand, the productivity of hours worked has increased due to the learning effect, which increases the demand for labor. The relative magnitude of the two effects determines whether the hours response is hump shaped.

Aggregate output, however, may show a hump-shaped response even if hours does not, since it is a function of the effective labor input $N_t = X_t H_t$. Given an initial increase in H_t the endogenous skill X_t is above its steady-state level in the second period. As long as the decay of hours H_t is slower than the increase of skills, output keeps rising and the model produces a hump-shaped response.

Figure 3 confirms the intuition. In Experiment 1 in the left column of the figure, we contrast the benchmark LBD model (Prior 1) with alternative values for the persistence of the transitory shock. We choose $\rho = 0$, 0.5, and 0.85. We use the posterior means from Table 3 for the other parameters. The graphs show that, while the model is capable of generating some persistence, even in the presence of i.i.d. shocks to B_t, it is very small and short lived. The LBD response function exhibits a hump only after ρ reaches 0.85. Note, however, that a quarterly autocorrelation of 0.85 is not an unreasonable value, as it implies a half life of about one year for the transitory shock. This value of ρ also delivers a much shorter-lived response of output that actually shows a marked improvement over the benchmark LBD model along this dimension. We now obtain a response that peaks after only two quarters. However, the overall time-series fit decreases as we change ρ. Compared to the posterior mode, the log-likelihood decreases by 6.0 for $\rho = 0.85$ and 131.7 for $\rho = 0.5$.

The second column, Experiment 2, shows the combined effect of changing μ, ϕ, and ρ. For comparison we also include the VAR posterior obtained from the data. With only small adjustments to the learning parameters and the persistence parameter ($\mu = 0.2$, $\phi = 0.7$, $\rho = 0.8$), the implied response of output to a transitory shock virtually coincides with its VAR counterpart. The log-likelihood differential is now 16.1.

5.5 Persistence: Evidence from Autocorrelations

Many univariate studies of output dynamics (e.g. Cochrane (1988)) find that output growth is positively autocorrelated over short horizons and only weakly autocorrelated over longer horizons. This finding is confirmed in our bivariate analysis. The first four rows of Table 5 summarize posterior mean predictions of the DSGE models and the VAR along with 95% confidence intervals. According to the VAR, both $corr(\Delta y_t, \Delta y_{t-1})$, and $corr(\Delta y_t, \Delta y_{t-2})$ are clearly positive. As pointed out by Cogley and Nason (1995), the standard RBC model predicts the autocorrelations of output growth to be essentially zero. The learning-by-doing mechanism, on the other hand, is able to generate positive autocorrelations. The first-order autocorrelation is 0.06 with Prior 1 and 0.10 with Prior 2. Higher-order autocorrelations decay toward zero. However, the autocorrelations calculated from the VAR seem to decay faster than the ones obtained from the LBD models.

The next three rows contain the formal evaluation statistics discussed in Section 3. The statistics are based on the first four lags of the output growth autocorrelations. Under both Prior 1 and Prior 2 the learning mechanism leads to improvements. The posterior expected L_q-loss for autocorrelations decreases from 0.116 (RBC) to 0.078 (LBD, Prior 1) and 0.058 (LBD, Prior 2). In addition to the quadratic losses we report the expected L_{χ^2} losses. A value close to one indicates that the model prediction lies far in the tails of the posterior density obtained from the VAR. Again, the LBD mechanism helps to reduce the statistic from 0.99 (RBC) to 0.97 (LBD, Prior 2) and 0.907 (LBD, Prior 2).

The bottom half of the table contains the results for the hours. Based on a lag-by-lag comparison and the joint L_q statistic the DSGE models correctly capture the high autocorrelation in the hours series. The fairly high values of the L_{χ^2} statistic indicate that the orientation of the VAR posterior contours makes the joint prediction of the four autocorrelations appear to lie far in the tails. Nevertheless, the relative ranking of the models is not affected.

5.6 Endogenous versus Exogenous Productivity

Finally, we investigate how the learning-by-doing mechanism affects the measured productivity. While the real-business-cycle theory builds on a pronounced cyclical productivity, either in measured total factor productivity (TFP) or labor productivity, many researchers suspect that the measured productivity may not reflect the true exogenous shifts in production technology. For unusually high productivity during expansions, they often point to cyclical utilization, (e.g., Burnside, Eichenbaum, and Rebelo (1995), or Bils and Cho (1994)), short-run increasing returns (e.g., Hall (1990)) or cyclical composition of industrial production (e.g., Basu and Fernald (1995)). Our model also predicts that TFP is likely to be higher during expansions due to the accumulated labor market experience. According to the LBD model measured TFP consists of two components: true exogenous productivity and the induced productivity due to learning-by-doing.

Figure 4 compares the posterior means of measured TFP and exogenous productivity from the LBD model based on Prior 1. Both series are graphed as percentage deviations from the trend. As expected, measured TFP does overstate the true productivity in expansions and understates in recessions. For example, TFP overstates the true productivity by about one percentage point in the late 1960's as the economy had accumulated a significant amount of human capital stock through learning-by-doing during an extended period of expansion. On the other hand, the decline of TFP is deeper than that of true productivity in the mid-1980's after the economy had experienced a series of recessions in the mid-1970's and early 1980's, both of which were associated with short hours of work.[9]

Nevertheless, the differences between measured TFP and the true exogenous productivity are fairly small. The estimated standard-error ratio of the endogenous versus the exogenous TFP component ranges from 8.8% to 12.2%. Thus, the discrepancy is not so dramatic as in Burnside, Eichenbaum, and Rebelo (1995). The reason is that in our model the endogenous productivity stems from the accumulated stock of human capital, which tends to move far less over time than the utilization rate of capital in Burnside *et al.* (1995). While we do not take a stand on the magnitude of other endogenous components of TFP, the contribution of learning-by-doing is modest.

6 Conclusion

Despite their popularity and wide application, standard real business-cycle models lack a satisfying internal-propagation mechanism. To generate such a mechanism, we augment the RBC model with learning-by-doing, where the current labor supply affects workers' future labor productivity.

Based on the individual-level panel data on wages and hours obtained from the PSID, we construct micro estimates for the learning-by-doing mechanism. These estimates are combined with time-series data on GDP growth and hours to perform a Bayesian analysis of the representative-agent model. We find that the LBD model fits the aggregate data much better than does the standard RBC model. Overall, the model, by and large, reproduces the impulse response in the VAR reasonably well. Yet the response of hours is delayed for about 2-3 quarters in the data, suggesting important frictions in the labor market. Nevertheless, we view learning-by-doing

[9]The results based on Prior 2 are very similar.

as an important propagation mechanism that can easily be built into more complicated DSGE models to improve their empirical performance. For instance, the mechanism could be easily incorporated in monetary models to amplify the real effect of monetary policy.

References

[1] Altug, S. and R. A. Miller (1998), "The Effect of Work Experience on Female Wages and Labour Supply," Review of Economic Studies 65, 45-85.

[2] Andolfatto, D. (1996), "Business Cycles and Labor Market Search," American Economic Review 86, 112-132.

[3] Basu S. and J. G. Fernald (1995), "Aggregate Productivity and the Productivity of Aggregates," mimeo, Hoover Institution.

[4] Baxter, M., and B. King (1991), "Production Externalities and Business Cycles," Discussion Paper 53, Federal Reserve Bank of Minneapolis.

[5] Bils, M. and J. Cho, (1994) "Cyclical Utilization," Journal of Monetary Economics 33, 319-354.

[6] Blanchard, O., and D. Quah (1989), "The Dynamic Effects of Aggregate Supply and Demand Disturbances," American Economic Review 79 (4), 655-73.

[7] Box, G. (1980), "Sampling and Bayes' Inference in Scientific Modelling and Robustness," Journal of the Royal Statistical Society A 143, 383-430.

[8] Burnside, C., and M. Eichenbaum, (1996) "Factor Hoarding and the Propagation of Business-Cycle Shocks," American Economic Review 86, 1154-74.

[9] Burnside, C., M. Eichenbaum, and S. Rebelo (1995) "Capital Utilization and Returns to Scale," NBER Macroeconomics Annual, 67-109.

[10] Cain, G. G. (1986) "The Economic Analysis of Labor Market Discrimination: A Survey," Handbook of Labor Economics 1, 693-785, edited by O. Ashenfelter and R. Layard, Elsevier Science Publisher.

[11] Cochrane, J. (1988) "How Big is the Random Walk in GNP?," Journal of Political Economy 96, 893-920.

[12] Cochrane, J. (1994) "Permanent and Transitory Components of GNP and Stock Prices," Quarterly Journal of Economics 109, 241-265.

[13] Cogley, T., and J. Nason (1995) "Output Dynamics in Real Business Cycle Models," American Economic Review, 85, 492-511.

[14] Cooper, R., and A. Johri (1998) "Learning by Doing and Aggregate Fluctuations," NBER Working Paper No. 6898.

[15] Cossa, R., J. Heckman, and L. Lochner (2000) "Wage Subsidies and Skill Formation: A Study of the Earned Income Tax Credit," Subsidies to Low Wage Employment, Edited by E. Phelps, Russell Sage Foundation, New York.

[16] DeJong, D., B. Ingram (1998) "The Cyclical Behavior of Skill Acquisition," Manuscript, University of Iowa.

[17] DeJong, D., B. Ingram and C. Whiteman (1996) "A Bayesian Approach to Calibration," Journal of Business Economics and Statistics, 14, 1-9.

[18] den Haan, W., G. Ramey, and J. Watson (2000) "Job Destruction and Propagation of Shocks," American Economic Review 90, 482-498.

[19] Doan, T., R. Litterman, and C. Sims (1984) "Forecasting and Conditional Projections Using Realistic Prior Distributions," Econometric Reviews 3, 1-100.

[20] Geweke, J. (1999a) "Computational Experiments and Reality," Manuscript, University of Iowa.

[21] Geweke, J. (1999b) "Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication," Econometric Reviews 18, 1-126.

[22] Ghez, G. R. and G. Becker (1975) "The Allocation of Time and Goods over the Life Cycle," New York, National Bureau of Economic Research, Columbia University Press.

[23] Hall, R. (1990) "Invariance Properties and Solow's Productivity Residual," in Growth, Productivity and Employment, P.A. Diamond (ed.) Cambridge, MA: MIT Press, 71-112.

[24] Hall, R. (1997) "Macroeconomic Fluctuations and the Allocation of Time," Journal of Labor Economics 15, s223-s250.

[25] Hall, R. (1999) "Amplification and Persistence of Employment Fluctuations through Labor Market Frictions," Handbook of Macroeconomics, ed., by J. Taylor and M. Woodford.

[26] Hansen, L.P. and J.J. Heckman (1996) "The Empirical Foundations of Calibration," Journal of Economic Perspectives 10 (1), 87-104.

[27] Horvath, M (1999) "Business Cycles and the Failure of Marginal Firms," Manuscript, Dartmouth Tuck School of Business.

[28] Jacobson, L., R. LaLonde, and D. Sullivan (1993) "Earnings Losses of Displaced Workers," American Economic Review 83, 685-709.

[29] Kim, K. and A.R. Pagan (1997): "Econometric Analysis of Calibrated Models." In: M.H. Pesaran and M. Wickens (eds.) Handbook of Applied Econometrics, Basil Blackwell, Oxford.

[30] King, R., C. Plosser, S. Rebelo (1988) "Production, Growth, and Business Cycles: I. Neoclassical Model," Journal of Monetary Economics 81 (4), 819-40.

[31] Parkin, M. (1988) "A Method for Determining Whether Parameters in Aggregate Models are Structural," Carnegie-Rochester Conference Series on Public Policy 29, 215-52.

[32] Perli, R., and P. Sakellaris (1998) "Human Capital Formation and Business Cycle Persistence," Journal of Monetary Economics, 42, 67-92.

[33] Prescott, E. (1986) "Theory Ahead of Business Cycle Measurement," Carnegie-Rochester Conference Series on Public Policy 25, 11-44.

[34] Pries, M. (1999) "Persistence of Employment Fluctuations: A Model of Recurring Job Loss," Manuscript, University of Maryland.

[35] Schorfheide, F. (2000) "Loss Function-Based Evaluation of DSGE Models," Journal of Applied Econometrics 15, 645-670.

[36] Shapiro, M. and M. Watson (1988) "Sources of Business Cycle Fluctuations," NBER Macroeconomics Annual 1988, 111-48.

[37] Topel, R. (1991) "Specific Capital, Mobility and Wages: Wages Rise with Seniority," Journal of Political Economy 99, 145-176.

[38] Topel, R. and M. Ward (1992) "Job Mobility and the Careers of Young Men," Quarterly Journal of Economics 107, 439-479.

[39] Willis, R. J. (1986) "Wage Determinants: A Survey and Reinterpretation of Human Capital Earnings Functions," Handbook of Labor Economics Vol. 1, pp. 525-602, edited by O. Ashenfelter and R. Layard, Elsevier Science Publisher.

A Micro-Level Estimation

A.1 Data Set

The PSID sample period is 1971-1992. The sample consists of heads of households and wives. Wage data for wives are available only since 1979. Wages are annual hourly earnings (annual labor incomes divided by annual hours). Nominal wages are deflated by the Consumer Price Index. The base year is 1983. Workers who worked less than 100 hours per year or whose hourly wage rate was below $1 (in 1983 dollars) are viewed as non-employed even though their employment status is reported as employed in the survey. We use workers who were employed in non-agricultural sectors and not self-employed. Descriptive statistics for the sample used in the estimation are reported in Table 6.

A.2 Estimation and Computation

Rewrite the quasi-differenced wage equation as

$$\begin{aligned}\tilde{w}_{i,\tau} &= w_{i,\tau} - \xi w_{i,\tau-1} - \tilde{\mu} h_{t-1} + \sum_{j=1}^{4} \tilde{\mu}(\tilde{\phi}^j - \xi \tilde{\phi}^{j+1}) h_{t-1-j} + \tilde{\mu}\xi\tilde{\phi}^5 h_{t-6} \\ &= \tilde{\zeta}_i + d_\tau' \tilde{\lambda}_d + v_{i,\tau}' \tilde{\lambda}_v + \epsilon_{i,\tau}^u\end{aligned}$$

or in matrix form for individual i

$$\tilde{W}_i = \iota_i \tilde{\zeta}_i + Z_i \tilde{\lambda} + E_i^u,$$

were ι_i is a column vector of ones, Z_i has rows $[d_\tau', v_{i,\tau}']$, and $\tilde{\lambda} = [\tilde{\lambda}_d', \tilde{\lambda}_v']'$. Moreover, define $\tilde{\theta} = [\tilde{\mu}, \tilde{\phi}, \xi, \sigma_u]'$. Thus, the conditional-likelihood function is

$$p(W|H, \tilde{\theta}, \tilde{\lambda}, \tilde{\zeta}_i) \propto \prod_{i=1}^{n} \sigma_u^{-1/2} \exp\left\{ -\frac{1}{2\sigma_u} (\tilde{W}_i - \iota_i\tilde{\zeta}_i - Z_i\tilde{\lambda})'(\tilde{W}_i - \iota_i\tilde{\zeta}_i - Z_i\tilde{\lambda}) \right\}.$$

This likelihood is combined with the Gaussian priors for $\tilde{\lambda}$ and $\tilde{\zeta}_i$ described in the text, which are specified conditional on $\tilde{\theta}$. Since both the log-likelihood and the log-prior density for $\tilde{\lambda}$ and $\tilde{\zeta}_i$ are quadratic in $\tilde{\lambda}$ and $\tilde{\zeta}_i$, it is straightforward to obtain a closed-form solution for the marginalized likelihood function:

$$p(W|H, \tilde{\theta}) = \int p(W|H, \tilde{\theta}, \tilde{\lambda}, \tilde{\zeta}) p(\tilde{\lambda}, \tilde{\zeta}|\tilde{\theta}) d\tilde{\lambda} d\tilde{\zeta}.$$

A Metropolis algorithm (see Section B.2.2) is used to generate posterior 100,000 draws from $p(\tilde{\theta}|W, H)$. The first 10,000 draws are discarded. The draws of $\tilde{\mu}$ and $\tilde{\phi}$ are converted into μ and ϕ. (See Equation (30.) Conditional on $\tilde{\theta}$, it is possible to compute posterior means and variances for $\tilde{\lambda}$. The unconditional means and variances reported in Table 1 were obtained by Monte Carlo averaging across posterior draws of $\tilde{\theta}$.

A.3 Conversion of Estimates from Annual to Quarterly

Unfortunately, an analytical solution for the conversion from annual to quarterly values for μ and ϕ is not available. Instead we approximate the quarterly values from annual estimates in the following way: After controlling for individual effects and the aggregate price of an effective unit of labor, the wage of an individual evolves according to his skill x accumulated from learning-by-doing. Because in the PSID wages are annual averages and hours are reported as cumulative annual hours, it is convenient to denote the time series of $\{x_t\}$ as $\{x_{\tau,q}\}$, where τ and q denote the year and the quarter of the observation. The log average wage in year τ, is a function of $\widetilde{x}_\tau = \frac{1}{4}\sum_{q=1,..,4} x_{\tau,q}$.[10] Assume $h_j = h$ for $j = t-s, ..., t$, where h is the steady-state level of log-hours. From our skill accumulation equation, $x_t = \phi x_{t-1} + \mu h_{t-1}$, the learning-by-doing effect survives at rate ϕ as

$$x_t - \frac{\mu h}{1-\phi} = \phi^s [x_{t-s} - \frac{\mu h}{1-\phi}].$$

Let $t = 1$ correspond to $\tau = 1$ and $q = 1$. The average hourly earning for year $\tau = 2$ can be written as

$$\widetilde{x}_2 = \frac{(x_{2,1} + x_{2,2} + x_{2,3} + x_{2,4})}{4}$$

[10] Note that $\frac{1}{4}\sum_{q=1}^{4} x_{\tau,q}$ can be interpreted as a log-linear approximation of the arithmetic average $\ln[\frac{1}{4}\sum_{q=1}^{4} e^{x_{\tau,q}}]$.

$$
\begin{aligned}
&= \frac{(x_5 + x_6 + x_7 + x_8)}{4} \\
&= \frac{\mu h}{1-\phi} + \frac{1}{4}\phi^4(\phi + \phi^2 + \phi^3 + \phi^4)(x_0 - \frac{\mu h}{1-\phi}) \\
&= \frac{\mu h}{1-\phi} + \frac{1}{4}\phi^4(x_1 + x_2 + x_3 + x_4 - 4\frac{\mu h}{1-\phi}) \\
&= \phi^4 \widetilde{x}_1 + \frac{1-\phi^4}{1-\phi}\mu h \\
&= \phi^4 \widetilde{x}_1 + \frac{1-\phi^4}{1-\phi}\mu \widetilde{h} + \text{constant.} \qquad (29)
\end{aligned}
$$

The last equation holds in the steady state because log annual hours are $\tilde{h} = h + \ln 4$. Equation (29) is equivalent to our regression of annual average wage on annual hours. This implies the following approximate relationships between the annual values of $\widetilde{\mu}$ and $\widetilde{\phi}$ and the quarterly underlying parameters μ and ϕ:

$$
\widetilde{\phi} = \phi^4 \quad \text{and} \quad \widetilde{\mu} = \frac{1-\phi^4}{1-\phi}\mu. \qquad (30)
$$

While the conversion formula is an approximation if h_t fluctuates around the steady state, we find that it is very accurate for our purpose. The approximation error is less than 3% when we simulate the x_t using a stochastic process that mimics the actual aggregate-hours and learning-by-doing processes in Equation (3).

B Aggregate Time-Series Analysis

B.1 Data Set

The following time series were extracted from DRI: gross domestic product (GDPQ), employed civilian labor force (LHEM), and civilian non-institutional population 20 years and older (PM20 and PF20). We defined population as $POP = PF20 + PM20$. From the BLS we obtained this series: average weekly hours, private non-agricultural establishments (EEU00500005). Prior to 1963 the BLS series is annual. We used these annual averages as monthly observations without further modification. Our monthly measure of hours worked is $H_t = \ln(EEU00500005 * LHEM \ / \ POP)$. We convert to quarterly frequency by simple averaging.

B.2 Vector Autoregression: Estimation and Computation

A fourth-order vector autoregression serves as a reference model:

$$y_t = C_0 + \sum_{h=1}^{4} C_h y_{t-h} + u_t \quad u_t \sim \mathcal{N}(0, \Sigma), \tag{31}$$

where y_t denotes a vector of GDP growth and hours worked. Define the $1 \times k$ vector $x_t = [1, y'_{t-1}, \ldots, y'_{t-4}]$ and the matrix of regression coefficients $C = [C_0|C_1|\ldots|C_4]'$. Let X denote the $T \times k$ matrix with rows x_t and Y a $T \times 2$ matrix with rows y'_t. Moreover, let $c = vec(C)$, where vec denotes the operator that vectorizes the columns of a matrix.

B.2.1 Prior

The Minnesota prior expresses the belief that the vector time series is well described as a collection of independent random walks. Consider equation i of the VAR model:

$$y_{i,t} = C_{i,0} + C_{i,tr} t + C_{i,1} y_{t-1} + \cdots + C_{i,4} y_{t-4} + u_t, \quad i = 1, 2. \tag{32}$$

Since $y_{1,t}$ is differenced output and the theory implies that hours are stationary, we choose the prior mean to be zero for all coefficients. The variance for $C_{i,0}$ is 100, which makes the prior diffuse. The variance of $C_{ij,l}$, $l = 1, \ldots, p$ is given by

$$var(C_{ij,l}) = \begin{cases} (\zeta/l)^2 & \text{if} \quad i = j \\ (\zeta\hat{\sigma}_i/l\hat{\sigma}_j)^2 & \text{if} \quad i \neq j, \end{cases} \tag{33}$$

where ζ is a hyperparameter. $\hat{\sigma}_i$ and $\hat{\sigma}_j$ are the OLS estimates of the error variance in equations i, j based on a short training sample. All prior covariances among different parameters are zero. The general structure of the prior for C is

$$vec(C) \sim \mathcal{N}\Big(vec(\bar{C}), V_c(\zeta)\Big). \tag{34}$$

To complete the specification we use an uninformative prior $p(\Sigma) \propto |\Sigma|^{-3/2}$ for the covariance matrix Σ. The prior for the hyperparameter ζ is uniform on the grid $\zeta \in \mathcal{Z} = \{\zeta(1), \zeta(2), \ldots, \zeta(J)\}$. We choose $\zeta_1 = 0.001$, $\zeta_J = 10$, $J = 20$, and $\ln \zeta_j$, equally spaced in the interval $[\ln \zeta_1, \ln \zeta_J]$.

B.2.2 Posterior

The Gibbs sampler is used to obtain 100,000 draws $(C^{(s)}, \Sigma^{(s)}, \zeta^{(s)})$, $s = 1, \dots, n_{sim}$ from the posterior distribution $p(C, \Sigma, \zeta | Y, \mathcal{M}_*)$ of the VAR parameters. The first 10,000 draws are discarded. The Gibbs sampler iterates over the following conditional distributions: $\zeta | Y, C, \Sigma$ (discrete), $C | Y, \Sigma, \zeta$ (multivariate normal), and $\Sigma | Y, C, \zeta$ (inverse Wishart). For each draw $(C^{(s)}, \Sigma^{(s)})$ we calculate the desired population moments and impulse response functions. This leads to draws from the posterior distribution of population characteristics $p(\varphi | data)$. Expected values are computed by Monte Carlo averaging.

We compute marginal data densities conditional on a training sample of 22 observations. The first 4 of the 22 observations are used to initialize lags. The conditional data density is proper and can be used to obtain posterior model probabilities. The marginal data density can be expressed as

$$p(Y_T | Y_0, \mathcal{M}_*) = \prod_{t=1}^{T} \int p(y_t | Y_{t-1}, Y_0, C, \Sigma, \zeta, \mathcal{M}_*) p(C, \Sigma, \zeta | Y_{t-1}, \mathcal{M}_*) d(C, \Sigma, \zeta), \tag{35}$$

where Y_0 denotes the training sample. For each t we approximate the integral by Monte Carlo integration. The Gibbs sampler is used to generate draws from $p(C, \Sigma, \zeta | Y_{t-1}, \mathcal{M}_*)$. At each step, we use 2000 burn-in draws that are discarded and 20,000 draws to approximate the integral.

B.3 DSGE Models: Estimation and Computation

Conditional on parameter values $\theta^{(i)}$, the likelihood function of the linearized DSGE models can be evaluated with the Kalman filter. A numerical-optimization routine is used to find the posterior mode. The inverse Hessian is calculated at the posterior mode. 100,000 draws from the posterior distributions of the DSGE model parameters are generated with a random-walk Metropolis algorithm. The first 10,000 draws are discarded. The inverse Hessian serves as a covariance matrix when we draw the proposed steps for the Metropolis algorithm. Geweke's (1999b) modified harmonic-mean estimator is used to approximate the marginal data densities of the DSGE models. Details of these computations are discussed in Schorfheide (2000).

Table 1: Panel Data Estimation Results

	Prior		Posterior	
	Mean	(S.E.)	Mean	(S.E.)
Learning-by-Doing Parameters (Annual)				
μ			0.3259	(0.0095)
ϕ			0.4070	(0.0244)
Learning-by-Doing Parameters (Quarterly)				
μ			0.1106	(0.0039)
ϕ			0.7973	(0.0122)
Additional Parameters				
ξ			0.5128	(0.0071)
σ_u			0.1982	(0.0011)
Ratio (AEP)	0.4000	(0.2000)	0.2621	(0.0245)
Peak (AEP)	55.000	(5.0000)	59.328	(2.9415)
Schooling (Years)	0.1000	(0.1000)	0.0867	(0.0040)
Female	-0.4000	(0.2000)	-0.2977	(0.0004)

Notes: Peak (AEP) refers to the peak of the age-earnings profile; *Ratio (AEP)* refers to the difference in log wages at the peak and at age 30. We are using an improper prior for annual μ, annual ϕ, ξ, and σ_u, which is proportional to σ_u^{-1}. The moments of μ, ϕ, ξ and σ_u are calculated from the output of the Metropolis algorithm. Conditional on μ, ϕ, ξ, and σ_u, we can calculate posterior moments for the remaining parameters analytically. The numbers reported in the table are obtained by Monte Carlo averaging of the exact conditional moments. The estimated simulation standard errors for the posterior moments are less than 1%.

Table 2: Prior Distributions for DSGE Model Parameters

Name	Range	Density	Mean	S.E.
		Learning-by-Doing Parameters: Prior 1		
μ	$I\!R^+$	Trunc. Bivar. Normal	0.111	(0.004)
ϕ	[0,1]		0.798	(0.012)
		Learning-by-Doing Parameters: Prior 2		
μ	$I\!R^+$	Trunc. Bivar. Normal	0.145	(0.078)
ϕ	[0,1]		0.664	(0.213)
		Additional Parameters		
α	[0,1]	Beta	0.660	(0.020)
β	[0,1]	Beta	0.993	(0.002)
γ	$I\!R$	Normal	0.005	(0.005)
δ	[0,1]	Beta	0.025	(0.005)
ν	[0,1]	Gamma	2.000	(0.500)
ρ	[0,1]	Beta	0.800	(0.100)
σ_a	$I\!R^+$	Inverse Gamma	N/A	N/A
σ_b	$I\!R^+$	Inverse Gamma	N/A	N/A

Notes: The parameters μ and ϕ appear only in the LBD model. The inverse Gamma priors are of the form $p(\sigma|\nu, s) \propto \sigma^{-\nu-1} e^{-\nu s^2/2\sigma^2}$, where $\nu = 1$ and $s = 0.015$. The first and second moments of this distribution do not exist. Its mode is equal to 0.010.

Table 3: Time-Series Estimation Results

	RBC		LBD (Prior 1)		LBD (Prior 2)	
	Mean	(S.E.)	Mean	(S.E.)	Mean	(S.E.)
			Learning-by-Doing Parameters			
μ			0.1106	(0.0039)	0.2013	(0.0746)
ϕ			0.7973	(0.0122)	0.4009	(0.1891)
			Additional Parameters			
α	0.6563	(0.0200)	0.6530	(0.0200)	0.6526	(0.0197)
β	0.9934	(0.0019)	0.9936	(0.0019)	0.9937	(0.0017)
γ	0.0040	(0.0009)	0.0040	(0.0009)	0.0040	(0.0009)
δ	0.0226	(0.0052)	0.0218	(0.0049)	0.0222	(0.0049)
ν	1.3088	(0.3196)	1.4853	(0.4512)	1.4189	(0.4052)
ρ	0.9442	(0.0255)	0.9371	(0.0255)	0.9387	(0.0248)
σ_a	0.0116	(0.0007)	0.0118	(0.0008)	0.0117	(0.0008)
σ_b	0.0089	(0.0013)	0.0088	(0.0017)	0.0086	(0.0018)

Notes: The posterior moments are calculated from the output of the Metropolis algorithm. The estimated simulation standard errors for the posterior moments are less than 1%.

Table 4: Prior and Posterior Model Probabilities

	RBC	LBD (Prior 1)	LBD (Prior 2)	VAR(4)
Prior Probabilities $\pi_{i,0}$	0.25	0.25	0.25	0.25
Data Density $\ln p(Y_T\|\mathcal{M}_i)$	1054.60	1055.92	1057.63	1082.72
Posterior Odds $\pi_{i,T}/\pi_{RBC,T}$	1.00	3.86	20.70	1.63E12
Posterior Probabilities $\pi_{i,T}$	0.6E-12	2.3E-12	12E-12	1.00

Notes: Marginal data densities for the DSGE models are computed by Geweke's (1999b) modified harmonic-mean estimator. The marginal data density of the VAR is computed via Monte-Carlo approximation of one-step-ahead predictive densities: $p(y_t|Y_{t-1},\mathcal{M}_0) \approx \frac{1}{n_{sim}}\sum_{s=1}^{n_{sim}} p(y_t|Y_{t-1},\theta^{(s)},\mathcal{M}_0)$, where $\theta^{(s)}$ is a draw from $p(\theta|Y_{t-1},\mathcal{M}_0)$. The posterior simulator for $p(\theta|Y_{t-1},\mathcal{M}_0)$ is described in the Appendix.

Table 5: Autocorrelation Statistics

	Lag	RBC	LBD (Prior 1)	LBD (Prior 2)	VAR
		Output Growth: $Corr(\Delta \ln GDP, \Delta \ln GDP(-j))$			
Post. Mean	1	0.008	0.060	0.101	$\underset{[0.151, 0.462]}{0.310}$
	2	0.007	0.049	0.046	$\underset{[0.016, 0.462]}{0.158}$
	3	0.007	0.040	0.023	$\underset{[-0.109, 0.170]}{0.029}$
	4	0.007	0.032	0.012	$\underset{[-0.158, 0.102]}{-0.027}$
L_q-risk	1-4	0.116	0.078	0.058	
$C_{\chi^2}(\hat{\varphi}\vert Y)$	1-4	16.01	11.96	8.177	
L_{χ^2}-risk	1-4	0.990	0.970	0.907	
		Hours: $Corr(\ln H, \ln H(-j))$			
Post. Mean	1	0.939	0.948	0.946	$\underset{[0.931, 0.983]}{0.957}$
	2	0.882	0.896	0.887	$\underset{[0.817, 0.952]}{0.886}$
	3	0.828	0.844	0.828	$\underset{[0.685, 0.921]}{0.800}$
	4	0.779	0.794	0.772	$\underset{[0.545, 0.884]}{0.710}$
L_q-risk	1-4	0.006	0.009	0.005	
$C_{\chi^2}(\hat{\varphi}_{q,i}\vert Y_T)$	1-4	42.73	25.19	20.46	
L_{χ^2}-risk	1-4	0.999	0.995	0.989	

Notes: As L_q-risk we only report $\tilde{R}_q(\hat{\varphi}|Y)$, defined in Equation (21). $C_{\chi^2}(\hat{\varphi}|Y)$ is defined in Equation (22). For the VAR we report 95% HPD confidence sets in brackets.

Table 6: Descriptive Statistics for PSID Subsample

Variable	Mean	Std.D.	No. of Obs.
Real Wage (in 1983 Dollars)	12.8	7.84	20145
Annual Hours of Work	2104.06	438.28	20145
Age	41.97	9.95	20145
Years of Schooling	13.22	2.43	20145
Gender (Female = 1)	0.31	0.46	20145

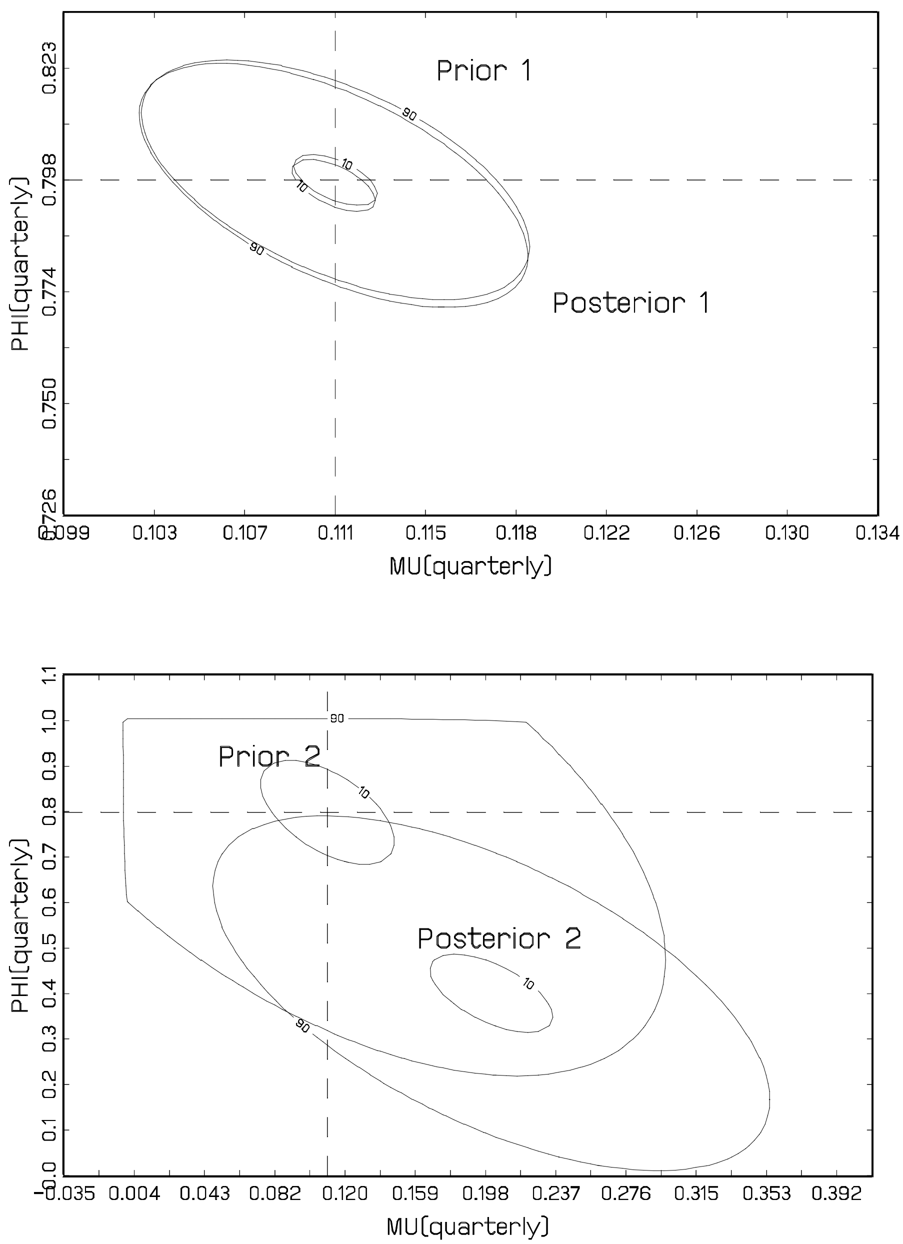

Figure 1: Confidence Sets (10% and 90%) for LBD Parameters

Notes: Prior 1 corresponds to the posterior distribution from the panel data analysis. Prior 2 is obtained by scaling the covariance matrix of Prior 1 and truncating the resulting distribution. The posteriors are based on the time-series estimation of the LBD model. Dashed lines intersect at modes of Prior 1 and 2.

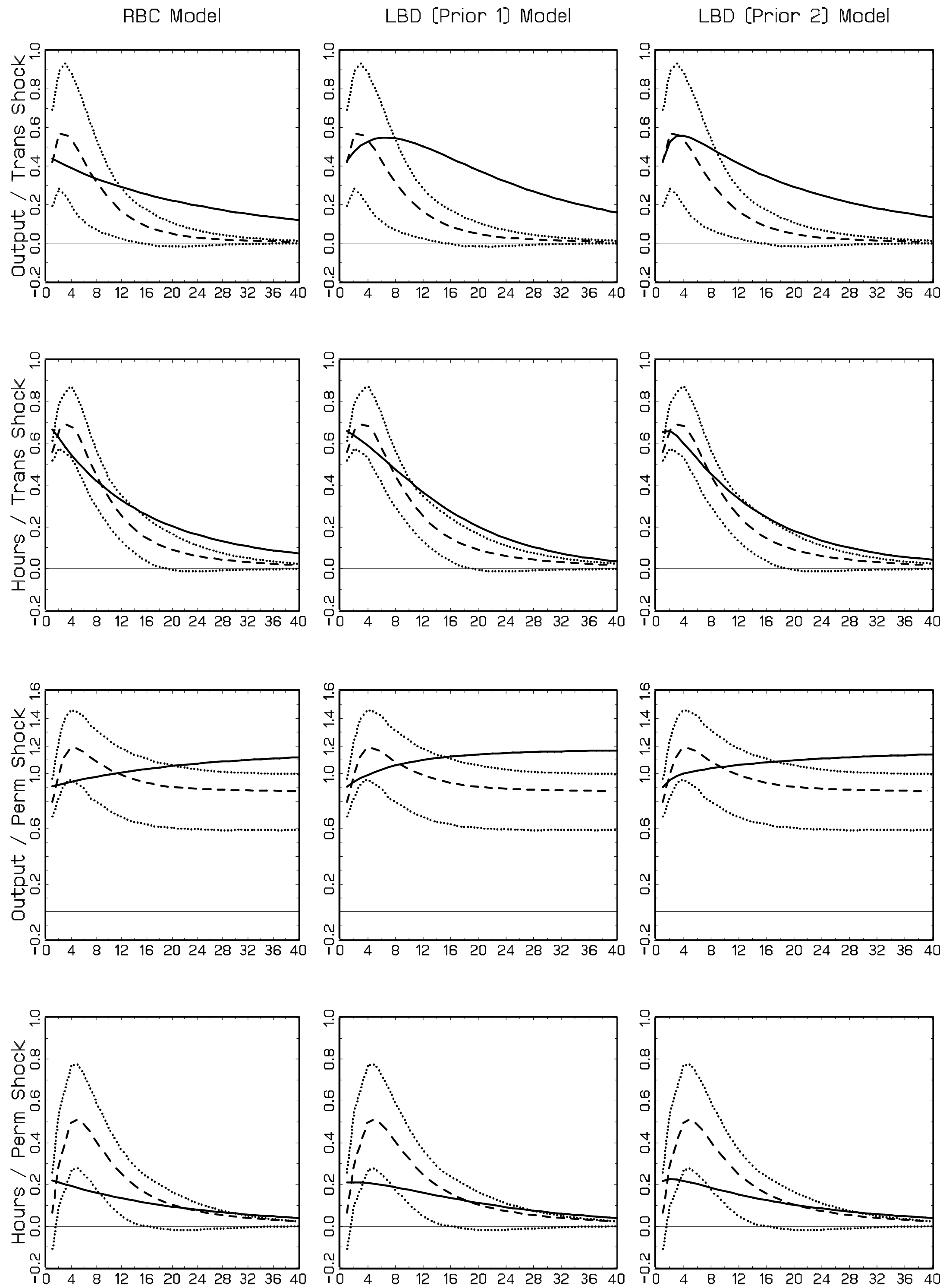

Figure 2: Impulse-Response Functions

Note: Solid lines are posterior means of DSGE models, *dashed lines* are posterior means of VAR, and *dotted lines* denote pointwise 75% HPD confidence intervals based on the VAR.

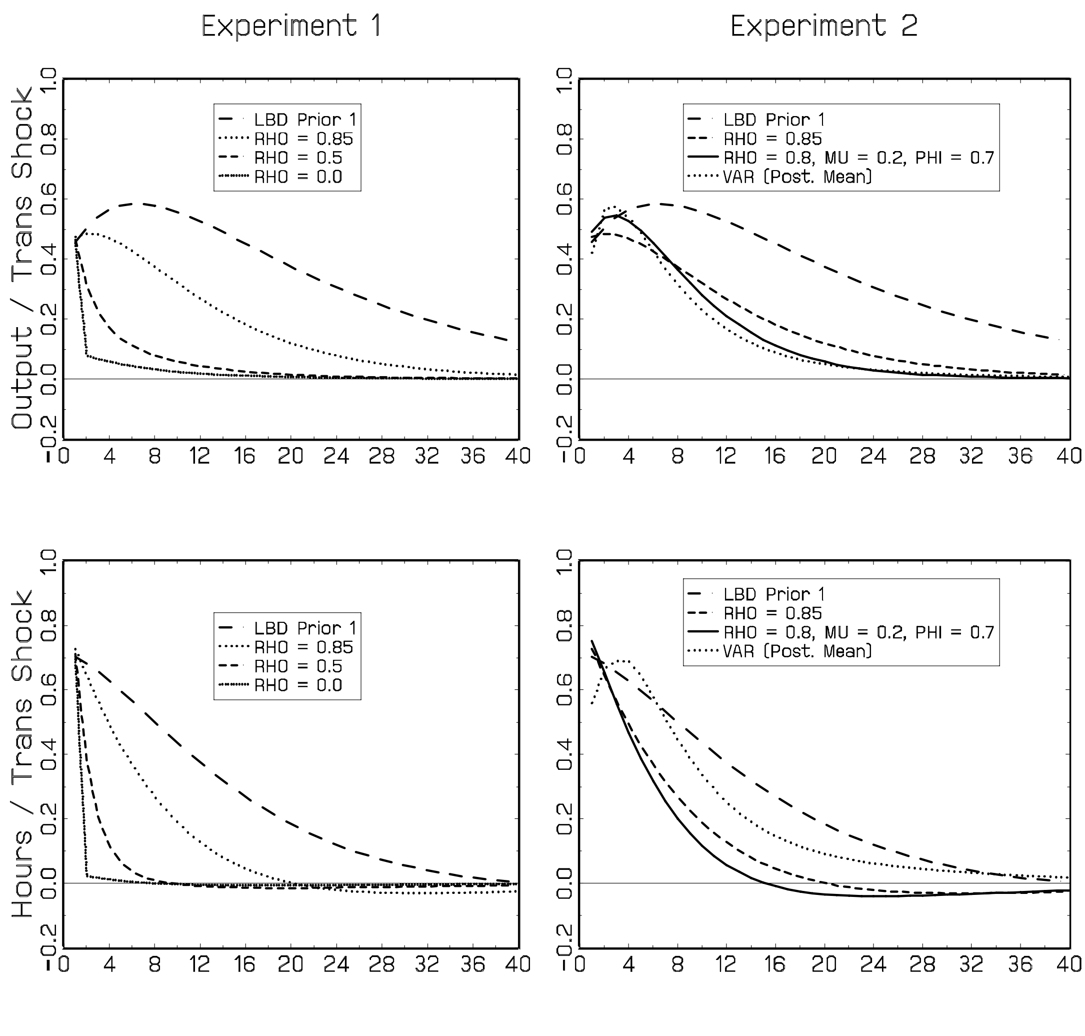

Figure 3: Exploring the Mechanism: Impulse Responses and Autocorrelation Functions

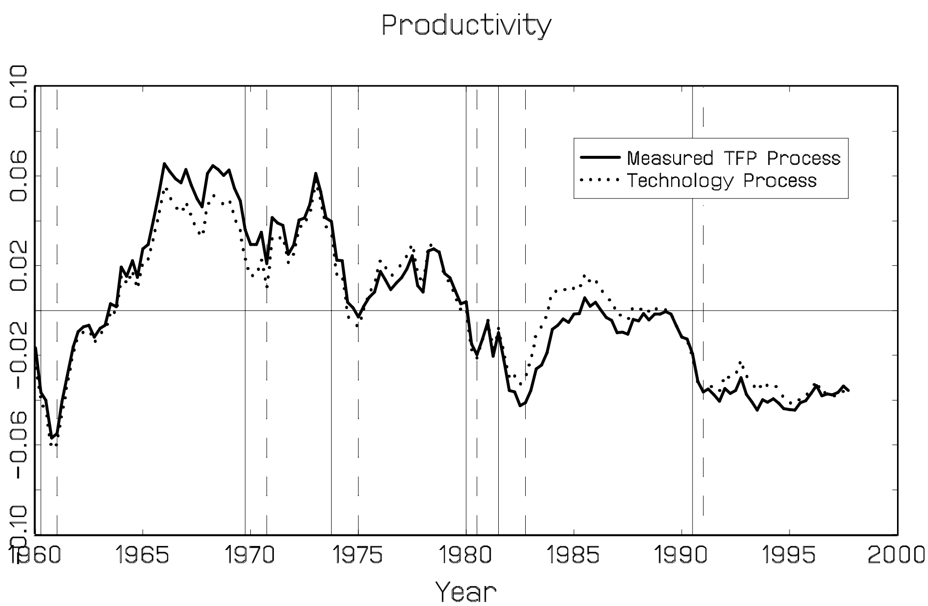

Figure 4: Exogenous Technology and Measured TFP

Notes: The technology processes are plotted in logs, detrended by the deterministic trend component γ^t. The graph depicts posterior means based on time-series estimation of the LBD model (Prior 1). *Solid vertical lines* correspond to business-cycle peaks, *dashed vertical lines* denote business cycle troughs (NBER business-cycle dating).