Persistence

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

University of Pennsylvania

January, 2000

Abstract

To generate persistence we augment the standard real business cycle (RBC) model with a "learning by doing" (LBD) mechanism, where current labor supply affects workers' future labor productivity. Our econometric analysis shows that the LBD model fits aggregate data much better than the standard RBC model. We calculate posterior odds for the structural models and formally show that the LBD model more closely mimics the autocorrelation and impulse response patterns that we found in a bivariate VAR analysis.

JEL CLASSIFICATION: C52, E32, J22

KEY WORDS: Aggregate Fluctuations, Econometric Model Evaluation, Learning by doing, Persistence

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

[†]Wharton School of Business, 3620 Locust Walk, Philadelphia, PA 19104-6302.

[‡]We thank Boyan Jovanovic, Narayana Kocherlakota, Donghoon Lee, James Nason, Victor Rios-Rull, Richard Rogerson, Randy Wright, and participants of NBER Summer Institute and Econometric Society Meeting for helpful comments.

1 Introduction

A well known shortcoming of the standard real business (RBC) model is its lack of an internal propagation mechanism. Aggregate output essentially traces out the movements of the exogenous technology process. This deficiency has been pointed out, among others, by Cogley and Nason (1995) and Rotemberg and Woodford (1996). While output growth is positively autocorrelated in the data, the model cannot generate any persistence in output growth from a random walk productivity process. It is also well known that GDP has an important trend-reverting component which is characterized by a hump-shaped response to a transitory shock, e.g., Blanchard and Quah (1989) and Cochrane (1994). However, the standard RBC model invariably generates a monotonic response of output in response to transitory shocks. In this paper, we introduce a fairly simple skill accumulation mechanism, which we will call "learning by doing", to overcome both deficiencies.

The standard model assumes that the worker's ability stays unchanged over time, despite the fluctuation of employment status and hours over the business cycle. However, extensive studies in labor economics have found important wage losses for displaced workers and a significant job-tenure effect in wage profiles. According to Ruhm (1991), based on the Panel Study of Income Dynamics (PSID), workers with displacement experience significant wage loss in the subsequent years of employment. At the same time displaced workers experience higher separation rates in subsequent jobs for years following displacement. On the other hand, wages of newly employed workers exhibit rapidly increasing profiles over time. These findings for individual workers fit into the broader picture of the behavior of aggregate economy. The business cycle associated with strongly procyclical hiring of new workers and countercyclical layoff of workers, indicates a systematic change in labor productivity of the work force.

We incorporate workers' on the job learning into a standard RBC model to account for the dependence of labor productivity on past work experience. This learning by doing (LBD) generates an internal propagation mechanism. The shifts in labor productivity outlive the exogenous technology shock. We refer to the modified RBC model as LBD model.

An econometric analysis is conducted to assess the empirical adequacy of the LBD specification and to quantify the improvements relative to the standard RBC model. A Bayesian approach lets us incorporate prior information on the parameters of the structural models. The priors are centered around values that are commonly used to calibrate RBC models, but they do not impose these values dogmatically.

To obtain a prior for the parameters of the learning by doing mechanism, we construct estimates from micro panel data. The Bayes estimation can be interpreted as follows: find values such that RBC and LBD model fit the data in a likelihood sense, without deviating too far from parameter values that are economically plausible and consistent with information from micro data sets.

The model comparison is based on the framework proposed by Schorfheide (1999). Instead of simply assessing the posterior odds of the LBD versus the RBC model, a vector autoregression (VAR) is used as a benchmark. We formally compare predictions of population moments and impulse response functions from the structural models to posterior estimates from the VAR. Both, the log-linearized structural models as well as the VAR provide linear moving average representations for aggregate data. The VAR representation, however, is less restrictive and therefore suitable to serve as a benchmark.

We find that the LBD model dominates the standard RBC model in many respects. First, introducing learning by doing significantly improves the statistical fit of the model. Approximate Bayes factors consistently favor the LBD specification, regardless of the prior distribution that we use. Second, the model with LBD successfully reproduces positive correlation in output growth even when exogenous technology follows random walk. Third, the impulse response function from the LBD clearly exhibits a hump-shaped response to a serially correlated transitory shocks as the current increase in hours leads to a subsequent increase in labor productivity.

A number of alternative explanations have been explored by other researchers. Cogley and Nason (1995) show that a straightforward modification of the basic model to allow for adjustment costs in capital fails to generate the required propagation of shocks. Hall (1999), Den Haan, Ramey and Watson (1997) and Pries (1999) examine the role of labor market frictions and job losses in generating persistent movements in unemployment and business cycle fluctuations. Perli and Sakellaris (1998) focus on intersectoral allocation of resources between market production and human capital accumulation.

Our work is closely related to Cooper and Johri (1998), who introduce the notion of learning by doing in the form of "organizational capital". Its level depends on past production. We introduce learning by doing through direct effects of past work experience on current labor productivity. Our approach has two advantages over the notion of organizational capital. First, we can measure the effects of learning by doing directly from panel data on wages and employment. Second, our modeling strategy avoids the issues of distinguishing between internal and external learning by doing, and determining which component of national income to match up with the contribution of LBD. The benefits to learning are incorporated in workers wages and thus will be included in the labor share of national income.

The paper is organized as follows. In Section 2 we augment the standard RBC model to allow for skill accumulation. Based on the panel data from the PSID, Section 3 discusses the evidence on the role of learning by doing. The econometric comparison of the RBC and the LBD model is conducted in Section 4. Section 5 summarizes our findings and discusses some avenues for future research.

2 A Stochastic Growth Model with Learning by Doing

The model economy is a variation of the standard stochastic growth model. Our main departure from the standard model is a learning by doing mechanism for labor. Workers' skill level varies over the business cycle along with their recent employment history.

There are several alternative ways of modelling the effects of skill accumulation in the production possibilities frontier. Cooper and Johri (1998) motivate learning by doing by introducing organizational capital as an additional input in production function, with the level of this organization capital depending on past production rates.

Introducing learning through labor productivity as we do has two important advantages. First, the model parameters can be directly infered from the ample panel data evidence on wages and work experience of individuals. Second, learning by doing usually creates an issue on whether it is internal or external or which component of national income is matched up with the contribution of learning by doing . Our modelling strategy avoids these issues since the benefit to learning are directly incorporated in workers wages and thus will be included in the labor share of national income.

2.1 Households

The representative household maximizes the expected discounted lifetime utility defined over consumption C_t and leisure $1 - H_t$ where H_t denotes hours spent at work:

$$U = E_t \left[\sum_{\tau=t}^{\infty} \beta^{\tau-t} [\log C_\tau + B_\tau \log(1 - H_\tau) \right] \tag{1}$$

where E_t denotes the expectation operator, conditional on information available at date t and β is the discount factor. B_t represents a stochastic preference shift.

The model economy is pertubated by two types of shocks: permanent shocks and transitory shocks. As it is common in the RBC literature, permanent shocks shift the production function. Transitory shocks are introduced through shifts in the marginal rate of substitution between goods and leisure which has been emphasized by Baxter and King (1991) and Hall (1997) as an important factor in explaining aggregate labor market fluctuations.[1]

Several studies in labor economics have documented the role of past labor supply on current wage determination. Two key findings are the existence of a job-tenure effect in wage profiles (Topel, 1991) and the significance of wage losses suffered by displaced workers, e.g., Ruhm (1991) and Jacobson, LaLonde and Sullivan (1993).

While most findings in this literature are based on binary employment status, we will introduce a continuous variable X_t that reflects past employment experience (which we identify with the skill level) in the representative agent economy. The basic findings above suggest the following parsimonious representation for the accumulation of skill, X_t:

$$\ln(X_t/X) = \phi \ln(X_{t-1}/X) + \mu \log(H_{t-1}/H), \quad 0 < \phi < 1, \mu > 0. \tag{2}$$

where variables without time subscript denote the stationary values.

In the present context equation (2) has a very simple interpretation. First it implies that an increase in the amount of hours worked in the current period contributes to an improvement in labor skills next period, with an elasticity of μ. Second, skill accumulation is persistent but not permanent. If hours worked fall below steady-state, skills decay over time at rate ϕ. Conveniently this model reduces to the standard stochastic growth model if $\mu = 0$, $X_t = X = 1/(1-\phi)$. The skill level X_t is interpreted as the effective unit of hours that the worker supplies. So the worker with skill X_t earns wage rate of $W_t(X_t) = W_t^* X_t$ where W_t^* denotes the market wage rate for efficiency unit of labor.

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

$$C_t + I_t = W_t(X_t)H_t + R_t K_t. \tag{3}$$

where I_t denotes capital goods expenditures. The accumulation of capital is described by the law of motion:

$$K_{t+1} = K_t + (1-\delta)K_t \tag{4}$$

[1]Having a preference shock is not critical for our purpose. To make the model consistent with the subsequent bivariate VAR analysis we only need a pair of shocks: a permanent shock and a transitory shock other than a technology shock.

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,

$$Y_t = K_t^{1-\alpha}(N_t Z_t)^{\alpha}. \tag{5}$$

Exogenous technological progress is denoted by Z_t. The labor input N_t consists of two components: an endogenous component that reflects job skill due to learning by doing, X_t, and hours worked, H_t. The effective number of hours is given by

$$N_t = X_t H_t \tag{6}$$

Profit maximization implies that firms solve the following problem

$$\max_{N_t, K_t} \quad (Z_t N_t)^{\alpha} K_t^{1-\alpha} - W_t^* N_t - R_t K_t \tag{7}$$

The optimality conditions for this problem are

$$W_t^* = \alpha Z_t^{\alpha} {N_t}^{\alpha-1} K_t^{1-\alpha} \tag{8}$$
$$R_t = (1-\alpha)(N_t Z_t)^{\alpha} K_t^{-\alpha} \tag{9}$$

3 Evidence on Skill Accumulation

The properties of the LBD model crucially depend on the parameterization of the skill accumulation process (2). First, we estimate the learning parameters ϕ and μ directly from the PSID for the period 1971-1992. Second, we compare our estimates to empirical evidence from a number of related microeconomic studies. The findings will be used to justify prior distributions of the LBD parameters for the time series analysis in Section 4.

3.1 Panel Data Estimation

Our model implies that the observed hourly log wage can be described as[2]:

$$w_{it} = w_t^* + \xi_{it} \tag{10}$$

[2]Unless otherwise noted, lower case letters are used to represent logs of upper case variables introduced earlier.

where ξ_{it} is individual specific term that affects the wage rate. The process ξ_{it} consists of the following components

$$\xi_{it} = x_{it} + \alpha_i + \beta t + \psi(L)\epsilon_{it} \tag{11}$$

where L denotes the temporal lag operator, and $\psi(L)$ is the polynomial $\sum_{j=0}^{\infty} \psi_j L^j$.

The process x_{it} is the micro level analogue to the aggregate skill $\ln X_t$. The idiosyncratic skill process evolves according to[3]

$$x_{it} = \phi x_{i,t-1} + \mu h_{i,t-1} \tag{12}$$

Here, $h_{i,t-1}$ denotes the average number of hours worked by individual i in period $t-1$. Since the number of hours that an individual can spend at the workplace is bounded, the skill level x_{it} is also bounded if $\phi < 1$. The skill of an individual who constantly works h hours converges to the maximum skill level $\mu h/(1-\phi)$.

Individual specific components are represented by the fixed effect α_i, capturing intrisic ability, and the process $\psi(L)\epsilon_{it}$, reflecting time varying idiosyncratic shocks such as the quality of match. These shocks reconcile the wage predicted by overall ability and occupational skill with the observed wage. For instance, an individual might receive an unusual low wage due to a poor match of qualification and job requirements. Alternatively, the wage could be unusually high because the employer overestimates his or her ability. At time t individuals i and their employers only know past and present ϵ_{it}'s.

The unobserved skill x_t can be eliminated by quasi-differencing of the wage equation. This leads to the dynamic panel data model

$$w_{it} = \phi w_{i,t-1} + \mu h_{i,t-1} + (1-\phi)\beta + (1-\phi L)w_t^* + (1-\phi)\alpha_i + (1-\phi L)\psi(L)\epsilon_{i,t} \tag{13}$$

The fixed effect $\alpha_i(1-\phi)$ can be removed by differencing (13) to get

$$\Delta w_{it} = \phi \Delta w_{i,t-1} + \mu \Delta h_{i,t-1} + (1-\phi L)(1-L)w_t^* + u_{i,t} \tag{14}$$

where

$$u_{i,t} = (1-\phi L)(1-L)\psi(L)\epsilon_{i,t} \tag{15}$$

The goal is to estimate ϕ and μ. We assume that the idiosyncratic disturbances ϵ_{it} are independent of aggregate process w_t^*. Since the market price of efficiency units

[3] Clearly this process only captures the role of *employment* skills (rather than *job specific* skills) on wages. To model the evolution of job specific skills it is necessary to assess how many of the previously acquired skills are transferable to the next job. This modeling exercise is beyond the scope of the present study.

of labor is unobserved, we introduce time dummies to capture w_t^* in the regression below.

$$\Delta w_{it} = \phi \Delta w_{i,t-1} + \mu \Delta h_{i,t-1} + \sum_{j=1}^{T} \delta_j \{t = j\} + u_{i,t} \tag{16}$$

where $\{t = j\}$ is the indicator function that is one if $t = j$ and zero otherwise.

Based on the interpretation of the idiosyncratic ϵ_{it} shocks given above, we will assume that their effects vanish after one period, i.e., one year. Thus, $\psi(L) = 1 + \psi_1 L$. The following orthogonality conditions are exploited to construct a GMM estimator for the skill accumulation parameters ϕ and μ

$$\mathbb{E}_{t-1}[u_{it} w_{i,t-h}] = 0, \quad \mathbb{E}_{t-1}[u_{it} h_{i,t-h}] = 0 \quad \text{for} \quad h = 4, 5 \tag{17}$$

If the time varying idiosyncratic skill process is a moving average of higher order, then our estimation procedure will generally be inconsistent. The direction of the estimation bias depends on the relative magnitude of ϕ the ψ_j's. Nevertheless, we think that our estimation procedure provides a suitable measurement of the order of magnitude of ϕ and μ.

The estimation results are summarized in Table 1. Since the estimates are obtained from annual panel data but our model is stated in terms of quarterly variables, we convert the numbers. Due to the non-linear conversion discussed in the Appendix, the standard errors for the quarterly estimates are smaller than for the annual estimates. Depending on the choice of the particular sample and the set of instruments, the quarterly point estimates range from 0.778 to 0.900 for ϕ, and 0.072 to 0.110 for μ.

3.2 Other Literature

A number of other studies confirms our findings about the existence of a fairly strong link between past work experience and current wages. Using a subset of PSID data 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. These numbers are even higher when total market experience is added to job-specific tenure.

The effects of job separations on wages are also well documented in the literature, although quantitative estimates are quite sensitive to the definition of separation being used. While focusing on very strict definitions of job separations usually leads to fairly high and persistent wage losses following a separation, broader definitions yield fairly negligible and much less persistent effects on wages.

Specification	Annual Estimates		Quarterly Estimates	
	$\tilde{\phi}$	$\tilde{\mu}$	ϕ	μ
All (I)	0.488 (0.162)	0.126 (0.109)	0.836 (0.069)	0.088 (0.154)
All (II)	0.515 (0.127)	0.133 (0.107)	0.847 (0.052)	0.087 (0.145)
All (III)	0.581 (0.116)	0.125 (0.112)	0.873 (0.044)	0.072 (0.142)
Men (I)	0.367 (0.169)	0.116 (0.172)	0.778 (0.090)	0.105 (0.251)
Men (II)	0.477 (0.143)	0.143 (0.179)	0.831 (0.062)	0.102 (0.240)
Men (III)	0.568 (0.131)	0.126 (0.190)	0.868 (0.050)	0.074 (0.237)
Women (I)	0.655 (0.343)	0.208 (0.169)	0.900 (0.118)	0.102 (0.269)
Women (II)	0.496 (0.238)	0.161 (0.144)	0.839 (0.101)	0.110 (0.220)
Women (III)	0.525 (0.216)	0.164 (0.148)	0.851 (0.088)	0.106 (0.214)

Table 1: Estimation Results for Learning by Doing Parameters. Standard errors are in parentheses. The conversion to quarterly estimates is explained in the appendix. Standard errors for quarterly observations are constructed via δ-method. Instruments (I): $w_{i,t-4}$, $w_{i,t-5}$, $l_{i,t-4}$, and $l_{i,t-5}$; Instruments (II): Instruments (I) and age; Instruments (III): Instruments (II) and years of schooling.

Thus, Jacobson, LaLonde and Sullivan (1993) focus on a sample of displaced workers only to find that wage losses are between 15% and 20% in the first year and a full recovery takes at least some 3 years, while Ruhm (1991), also looking at displacements, finds that workers wages may actually never recover completely from the initial, similar, wage losses. In the context of our model this would suggest annual values for μ and ϕ around 0.15 and 1, respectively. On the other extreme, Topel and Ward (1992) actually document that for young workers wages actually rise somewhat following job changes, suggesting that μ and ϕ should both be close to 0.

We account for these differences by imposing different priors on the distribution of the learning parameters in the analysis below. In particular we will consider two prior distributions based on our panel estimates and one prior distribution that is centered at $\phi = 0.5$. The latter can be regarded as a midpoint "estimate" based on the studies discussed in this subsection.

4 Econometric Model Evaluation

In this section we evaluate the empirical adequacy of the LBD model and compare it to the RBC model. Three specifications of the LBD model are considered that differ with respect to the prior distribution for the persistence parameter ϕ of the skill process. Based on the log-linearized structural models we derive a joint probability distribution for macroeconomic aggregates. Since the models contain two exogenous processes, Z_t and B_t, marginal distributions for 2×1 vectors of macroeconomic variables are non-singular. To keep our analysis comparable to that of earlier bivariate studies we fit the models to aggregate output growth and hours data.

4.1 Methodology

The models are evaluated based on their ability to reproduce regular features of aggregate output and employment data. For several reasons, we will adopt a Bayesian approach. First, it is straightforward to incorporate additional information into the parameter estimation. This information stems from two sources: calibration exercises as they are typically conducted for the parametrization of RBC models, and the learning by doing parameter estimates discussed in Section 3. Second, Bayesian procedures are ideal for model comparisons and the assessment of their relative statistical fit, while frequentist procedures are suitable to assess the absolute adequacy of one particular model at a time. A detailed argument along these lines can be found in Box (1980).

The parsimoneous and stylized nature of the model economies is a potential source for misspecification. For this reason, we do not limit our analysis to the calculation of posterior probabilities for the RBC and LBD model. Instead we employ the econometric framework proposed in Schorfheide (1999). To account for the potential misspecification of the structural models, we consider a vector autoregression as a reference model. Both the structural models as well as the VAR provide a linear moving average representation for aggregate output growth and hours. The VAR, however, is much less restrictive. Moreover, it is a popular tool in empirical macroeconomics.

The VAR is helpful for the model evaluation in two respects. First, the posterior odds of the structural models versus the VAR provide a measure of the overall statistical fit of the RBC and LBD model within the class of linear models. If the structural models due to their parsimony dominate the VAR then posterior odds are useful for comparing RBC and LBD model. On the other hand, if the statistical fit of the structural 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 proceeds by comparing the predictions of the structural models and the posterior VAR estimates with respect to population characteristics that a researcher is interested in. The structural model that best matches the posterior estimates wins the comparison. A formal description of the procedure is provided in the next subsection.

4.2 The Model Evaluation Procedure

The structural models are denoted by $\mathcal{M}_i$, $i = 1, \ldots, k$. The reference model is denoted by $\mathcal{M}_*$. $\theta^{(i)}$, $i = *, 1, \ldots, k$ is the parameter vector for model i. Let $\pi_{i,0}$ and $\pi_{i,T}$ be prior and posterior model probabilities, respectively. The $m \times 1$ vector φ contains population characteristics, such as truncated autocorrelation and impulse response functions. The posterior of φ conditional on a model $\mathcal{M}_i$ is denoted by the density $p(\varphi|data, \mathcal{M}_i)$. The econometric evaluation consists of the following steps.

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

$$\pi_{i,T} = \frac{\pi_{i,0} p(data|\mathcal{M}_i)}{\sum_{i=*,1,\ldots,k} \pi_{i,0} p(data|\mathcal{M}_i)} \tag{18}$$

where $p(data|\mathcal{M}_i)$ is the marginal data density[4].

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

$$p(\varphi|data) = \sum_{i=*,1,\ldots,k} \pi_{i,T} p(\varphi|data, \mathcal{M}_i) \tag{19}$$

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

$$p(\varphi|data) \approx p(\varphi|data, \mathcal{M}_*) \tag{20}$$

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 will examine the expected loss associated with $\hat{\varphi}_{i,b}$ under the posterior distribution of φ conditional on the VAR:

$$R(\hat{\varphi}_{i,b}|data, \mathcal{M}_*) = \int L(\varphi, \hat{\varphi}_{i,b}) p(\varphi|data, \mathcal{M}_*) \tag{21}$$

This posterior prediction risk $R(\hat{\varphi}_{i,b}|data, \mathcal{M}_*)$ provides an absolute measure of fit. The differential across structural models provides a relative measure of fit that allows model comparisons. Since $\hat{\varphi}_i$ is a function of the model parameters $\theta^{(i)}$ one can obtain loss function parameter estimates $\hat{\theta}_l^{(i)}$ by minimizing $R(\hat{\varphi}[\theta^{(i)}]|\text{data}, \mathcal{M}_*)$ with respect to $\theta^{(i)}$. These estimates provide a lower bound for the posterior risk attainable through a particular structural model.

We consider two loss functions to evaluate the deviations of actual model predictions from population characteristics. The quadratic loss function

$$L_q(\varphi, \hat{\varphi}) = (\varphi - \hat{\varphi})' W (\varphi - \hat{\varphi}) \tag{22}$$

with $m \times m$ weighting matrix W penalizes the distance between model predictions and posterior mean predictions of the reference model. Since the ranking of predictions $\hat{\varphi}$ depends only on the weighted distance between $\hat{\varphi}$ and the posterior mean of φ, an informal comparison of the two quantities can be interpreted as an evaluation under a quadratic loss function.

The second loss functions penalizes predictions that fall far into the tails of the overall posterior distribution $p(\varphi|data)$ of the population characteristics. Let $\bar{\varphi}$ be the posterior mean and V_φ the posterior covariance matrix. Define

$$L_{\chi^2}(\varphi, \hat{\varphi}) = \mathcal{I}\left\{ (\varphi - \bar{\varphi})' V_\varphi^{-1} (\varphi - \bar{\varphi}) < (\hat{\varphi} - \bar{\varphi})' V_\varphi^{-1} (\hat{\varphi} - \bar{\varphi}) \right\} \tag{23}$$

[4]The marginal data density is defined as $p(data|\mathcal{M}_i) = \int p(data|\theta^{(i)}, \mathcal{M}_i) p(\theta^{(i)}|\mathcal{M}_i) d\theta^{(i)}$, where $p(data|\theta^{(i)}, \mathcal{M}_i)$ is the likelihood function for model i and $p(\theta^{(i)}|\mathcal{M}_i)$ is the prior distribution of $\theta^{(i)}$

The expected L_{χ^2} loss is similar to a p-value if the posterior density is well approximated by a unimodal Gaussian density. However, its interpretation is different from traditional p-values. Most importantly, the expected L_{χ^2} loss can be used to formally rank misspecified models based on how far their predictions lie in the tails of the overall posterior distribution of population characteristics.

4.3 Results

All structural models as well as the VAR 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. Data definitions are provided in the Appendix. The RBC and LBD model are log-linearized and solved by standard methods. Conditional on parameter values $\theta^{(i)}$ the likelihood function can be evaluated with the Kalman filter. The likelihood is combined with an informative prior distribution described below. A numerical optimization routine is used to compute posterior modes. A fourth order vector autoregression serves as reference model $\mathcal{M}_*$. A version of the Minnesota prior (Doan, Litterman, and Sims, 1984) is used for the VAR coefficients. Draws from the posterior distribution of the VAR parameters are obtained by Gibbs sampling. For each of these draws we calculate the desired population moments and impulse response functions. This leads to draws from the posterior distribution of population characteristics $p(\varphi|data, \mathcal{M}_*)$. These draws can be used to determine posterior expected losses of model predictions. Details are provided in the Appendix.

4.3.1 Prior Distributions for the Structural Parameters

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 by matching steady state characteristics of the models to first moments of time series data, from micro econometric studies with cross sectional data, or by pure introspection. This feature of the calibration approach can be interpreted as a prior distribution that concentrates on a single point of the parameter space. Following the literature on Bayesian analysis of the models we relax the “tightness” of the prior and consider non-degenerate distributions. This prior is later combined with the likelihood function to obtain a posterior distribution. Intuitively the Bayes estimation can be interpreted as searching for parameter values such that the models fit the data, in a likelihood sense, without deviating too far from economically sensible values.

Name	Range	Density	Mean	Std.Error
α	[0,1]	Beta	0.660	(0.020)
β	[0,1]	fixed	0.990	N/A
γ	$I\!R$	Gaussian	0.005	(0.002)
δ	[0,1]	fixed	0.025	N/A
L_m	[0,1]	Beta	0.330	(0.020)
ρ	[0,1]	Beta	0.900	(0.100)
ψ	$I\!R$	Gaussian	0.000	(1000)
σ_ϵ	$I\!R^+$	InvGamma	∞	(∞)
σ_η	$I\!R^+$	InvGamma	∞	(∞)
Learning-by-Doing Parameters: Prior 1				
μ	$I\!R^+$	Gamma	0.090	(0.150)
ϕ	[0,1]	Beta	0.840	(0.060)
Learning-by-Doing Parameters: Prior 2				
μ	$I\!R^+$	Gamma	0.090	(0.150)
ϕ	[0,1]	Beta	0.840	(0.010)
Learning-by-Doing Parameters: Prior 3				
μ	$I\!R^+$	Gamma	0.090	(0.150)
ϕ	[0,1]	Beta	0.500	(0.060)

Table 2: Prior Distribution for the Parameters of the DSGE Models. The parameters μ and ϕ appear only in the LBD model.

The marginal prior densities that are used in the empirical analysis are summarized in Table 2. The shapes of these densities are chosen to match the domain of the structural parameters.

Naturally the skill accumulation parameters μ and ϕ are the focus of our analysis. To accommodate the different empirical estimates discussed earlier we will choose three different priors for them. Prior 1 is based on the panel data estimates presented in Section 3. The prior mean of ϕ is 0.84 which is approximately an average of the point estimates that we obtained for the "All" and "Men" samples. The prior standard error is 0.06 which somewhat understates the actual parameter uncertainty based on the Panel estimates. The prior mean for μ is 0.09 with a large standard error of 0.15. In the second prior we reduce the standard error of ϕ to 0.01, which pulls the posterior more strongly toward 0.84. The third prior for ϕ centered around 0.5.

The remaining prior means are calibrated to match the values that are commonly used in the RBC literature. We fixed the discount factor $\beta = 0.99$ and the depreciation rate $\delta = 0.025$ a priori. The priors for the labor share α and the steady state hours at work L are centered at 0.66 and 0.33, respectively. Since total hours are normalized to one in the structural models, we introduce a parameter ψ that corresponds to the log steady state value of actual hours worked. We use a diffuse prior for ψ, with mean zero and a large variance. The a priori expected value of ρ, the autoregressive coefficient for the preference process B_t is 0.9.

4.3.2 Posterior Distributions and Model Probabilities

Posterior mode estimates of the structural model parameters are summarized in Table 3. The RBC model is labeled $\mathcal{M}_1$ and the LBD model with Prior $j = 1, 2, 3$ is denoted by $\mathcal{M}_{2(j)}$. Most interesting for our analysis are the estimates of the learning parameters ϕ and μ. Under Prior 1 the posterior mode is equal to 0.99 which implies that the skill process is highly persistent.[5] The posterior density appears to be dominated by the contribution of the likelihood function. Under priors 2 and 3, however, the posterior mode of ϕ is very close to the respective prior means. The prior is very informative relative to the likelihood and the data does not lead to a substantial revision of the parameter values.

[5] Chang and Kwark (1999) report that aggregate hours exhibits extreme persistence in the data—one cannot reject the existence of a unit root. This finding is incompatible with the standard model driven by persistent productivity shocks because hour is little affected by permanent shifts in technology as income effect and substitution effect offset each other. It seems that the learning by doing parameter reflects this persistence in hours series in the data.

The likelihood values and marginal data densities, summarized in Table 4, shed more light on the interpretation of the parameter estimates. Most important in the table are the marginal data densities for the structural models and the reference autoregression. The marginal densities $p(Y_T|Y_0, \mathcal{M}_i)$ can be interpreted as maximum likelihood values, penalized for the dimensionality of the various models and adjusted for the effect of the prior distribution (see Appendix for details). While the reference model clearly outperforms the structural models and has essentially posterior probability one, it is nevertheless informative to look at the likelihood fit of the structural models.

There is a significant discrepancy between the value of the likelihood functions at their respective maxima and at the posterior modes, indicating the important role our priors on the structural parameters. The posterior mode is often quite similar to the prior mean. Unfortunately the parameter values that maximize the likelihood function are in many instances quite different from the priors. For instance, the likelihood estimates of the labor share parameter α are all greater than 0.9, a value that is economically implausible. Forcing the parameters to be in a "plausible" region of the parameter space however, has a cost in terms of likelihood fit.

Among the four structural specifications, the learning model with Prior 1 attains the best fit. The RMSE statistics suggest that all four specifications fit the output growth series about equally well. However, there are substantial differences with respect to the hours series. Intuitively the introduction of skill accumulation allows for a much smoother and highly persistent behavior in hours worked, particularly when the coefficient ϕ is high.

Regardless of the prior we consistently find that introducing learning significantly improves the fit over the standard the RBC model. Converted onto a posterior odds scale the marginal data density values imply that the odds of $\mathcal{M}_1$ versus $\mathcal{M}_{2(2)}$ or $\mathcal{M}_1$ versus $\mathcal{M}_{2(3)}$ are of order 1 to 4000 or worse[6].

Table about here

Table 3: Posterior Distribution of Model Parameters.

Table about here

Table 4: Marginal Data Densities.

[6]There is little discrimination between the specifications $\mathcal{M}_{2(2)}$ and $\mathcal{M}_{2(3)}$.

	$\mathcal{M}_1$: RBC		$\mathcal{M}_{2(1)}$: LBD Prior 1		$\mathcal{M}_{2(2)}$: LBD Prior 2		$\mathcal{M}_{2(3)}$: LBD Prior 3	
	mode	(stdd)	mode	(stdd)	mode	(stdd)	mode	(stdd)
α	0.7093	(0.0183)	0.6838	(0.0191)	0.7186	(0.0181)	0.7111	(0.0187)
β	0.9900		0.9900		0.9900		0.9900	
γ	0.0046	(0.0006)	0.0065	(0.0010)	0.0050	(0.0007)	0.0048	(0.0068)
δ	0.0250		0.0250		0.0250		0.0250	
L_m	0.4320	(0.0096)	0.3430	(0.0203)	0.3908	(0.0191)	0.4057	(0.0183)
ρ	0.9447	(0.0257)	0.9286	(0.0221)	0.9305	(0.0228)	0.9286	(0.0262)
ψ_1	3.1541	(0.0110)	3.1559	(0.0174)	3.1559	(0.0308)	3.1556	(0.0633)
μ			0.0724	(0.0056)	0.1248	(0.0100)	0.2585	(0.0610)
ϕ			0.9900	(0.0080)	0.8429	(0.0091)	0.4937	(0.0089)
σ_a	0.0110	(0.0006)	0.0116	(0.0008)	0.0113	(0.0007)	0.0109	(0.0007)
σ_b	0.0113	(0.0007)	0.0097	(0.0007)	0.0103	(0.0007)	0.0104	(0.0008)

Table 3: Posterior Mode for the Parameters of the DSGE Models. The standard errors refer to the diagonal elements of the inverse Hessian of the posterior density, evaluated at the posterior mode. The parameters β, and δ were fixed during the estimation.

	$\mathcal{M}_1$: RBC	$\mathcal{M}_{2(1)}$: LBD 1	$\mathcal{M}_{2(2)}$: LBD 2	$\mathcal{M}_{2(3)}$: LBD 3	VAR(4)
Likelihood Prior at Mode	1021.53	1062.51	1039.70	1036.02	
Likelihood at Mode	1024.37	1057.12	1032.26	1033.02	
Maximum Likelihood	1076.54	1076.84	1076.84	1076.84	1112.39
Marginal Density (MC Approx)	N/A	N/A	N/A	N/A	1080.00
Marginal Density (Laplace Approx)	991.669	1022.37	1001.13	1000.02	N/A
In-Sample RMSE Δy_t	0.00951	0.00957	0.00981	0.00949	0.00852
In-Sample RMSE L_m	0.00810	0.00721	0.00785	0.00801	0.00542

Table 4: Approximations to the log marginal data densities of structural models and the reference model. "Mode" refers to the mode of the posterior density. Likelihood Prior is $\ln p(Y_T|Y_0,\tilde{\theta}_i,\mathcal{M}_i)p(\tilde{\theta}_i|\mathcal{M}_i)$. Likelihood is $\ln p(Y_T|Y_0,\tilde{\theta}_i,\mathcal{M}_i)$. Maximum Likelihood is $\ln p(Y_T|Y_0,\tilde{\theta}_i,\mathcal{M}_i)$, Marginal density is $\ln p(Y_T|Y_0,\mathcal{M}_i)$. In-sample RMSE's are calculated at the posterior mode estimate for the structural models and at the maximum likelihood estimate for the VAR.

4.3.3 Persistence: Evidence from Autocorrelations

The learning feature was added to the real business cycle model in order to generate more persistence in aggregate output. 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. Figure 1 shows marginal posterior densities for autocorrelations of order one to four. Both $corr(\Delta y_t, \Delta y_{t-1})$, and $corr(\Delta y_t, \Delta y_{t-2})$ are clearly positive. The horizontal bars in the figure mark the posterior mode predictions of the various models. As pointed out by Cogley and Nason (1995), the standardard RBC model predicts the autocorrelations of output growth to be essentially zero. The learning mechanism, on the other hand, is able to generate positive autocorrelations. An informal inspection of the plots indicates that Prior 3 provides the best match a posteriori. However, the autocorrelations calculated from the VAR seem to decay faster than the ones obtained from the LBD models.

In Table 5 we report posterior expected L_{χ^2} losses. A value close to one indicates that the model prediction lies far in the tails of the posterior density. A value of zero means that the model prediction coincides with the posterior mean under the reference model $\mathcal{M}_*$. The expected losses are calculated jointly for lags 1 to 4 and lags 1 to 8. While the improvements under Priors 1 and 2 are moderate, specification $\mathcal{M}_{2(3)}$ fits substantially better than the standard RBC model and is able to capture the persistence of output growth. We also report L_{χ^2} losses for the autocorrelation of hours, for which the improvements through the learning mechanism are smaller.

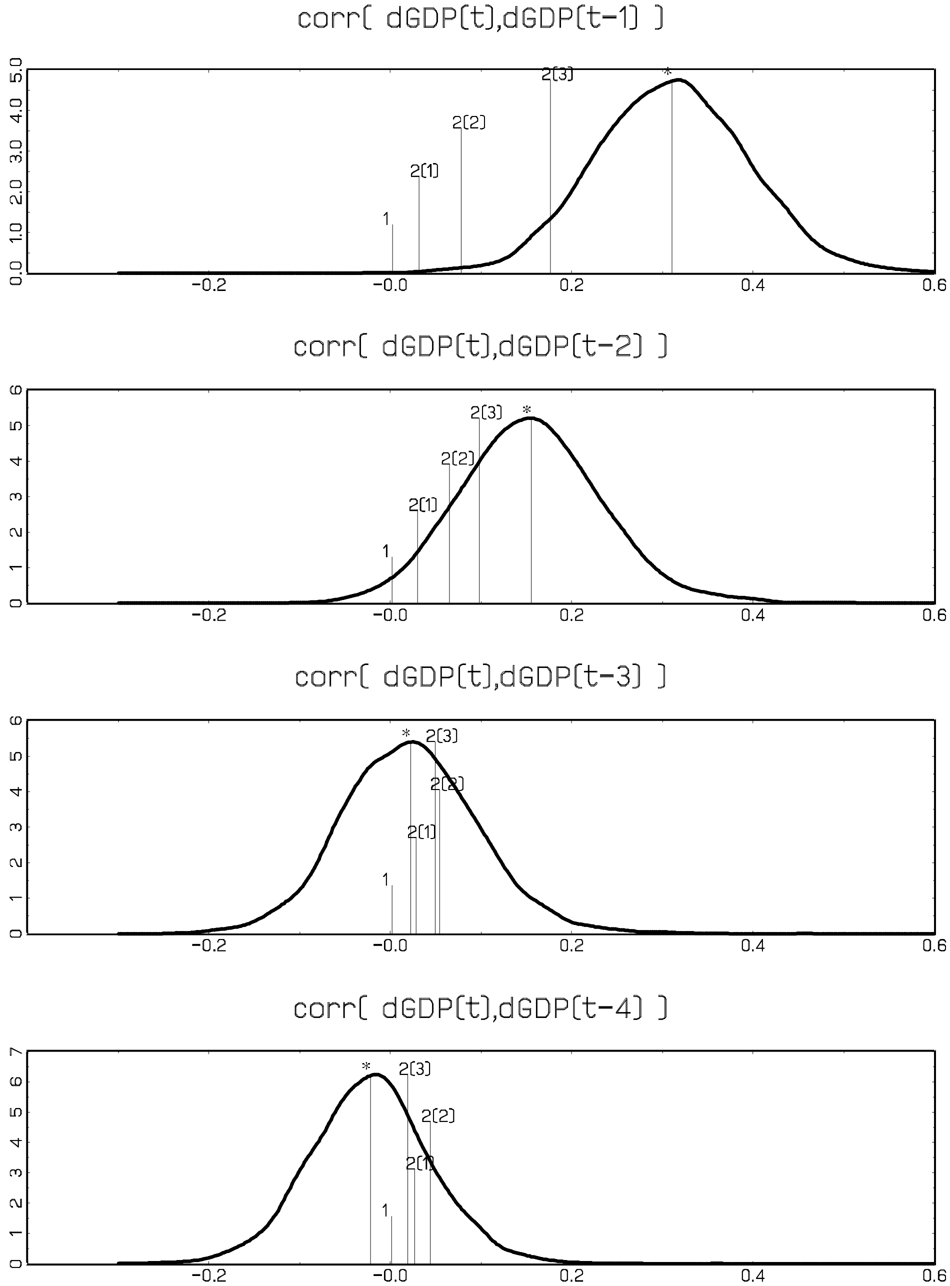

Figure 1: Marginal Posterior Densities for Autocorrelations of Output Growth. Kernel Density Estimates Based on Draws from Posterior Distribution. Horizontal Bars Refer to DSGE Model Predictions (Posterior Mode): 1 is RBC, 2(i) is LBD with Prior i.

Moment	Lags	$\mathcal{M}_1$	$\mathcal{M}_{2(1)}$	$\mathcal{M}_{2(2)}$	$\mathcal{M}_{2(3)}$
$corr(\Delta y_t, \Delta y_{t-h})$	1 - 4	0.9862	0.9790	0.9564	0.6170
	1 - 8	0.9186	0.9112	0.8774	0.5144
$corr(L_t, L_{t-h})$	1 - 4	0.9984	0.9982	0.9922	0.9680
	1 - 8	0.9866	0.9844	0.9600	0.8966

Table 5: Joint L_χ Prediction Losses for $corr(\Delta y_t, \Delta y_{t-h})$ and $corr(L_t, L_{t-h})$.

4.3.4 Impulse Response Dynamics

The dynamic behavior of time series models with an autoregressive structure can be summarized by impulse response functions (IRF). The four structural model specifications are driven by a random walk technology process and a stationary preference process. The innovations to the technology process have a permanent effect on output, whereas the innovations of the preference process have a transitory effect. Blanchard and Quah's (1989) method is used to identify responses to transitory and permanent shocks in the vector autoregression. We then assess the discrepancy between the model IRFs and VAR impulse response functions.

Before it is possible to compare IRFs we have to discuss their normalization. Since we have variance estimates for the structural shocks of the models and the VAR innovations it is possible to compare responses to one standard deviation shocks. This approach can be viewed as probabilistic normalization. The computation of posterior odds showed that the probabilistic structure of the models is rejected against the reference model. Therefore, we will choose a normalization that allows us to compare responses to shocks that lead to the same long-run or short-run effect. The models' technology shocks are normalized such that the long-run response of output equals the long-run posterior mean response to a one-standard deviation shock in the VAR. The preference shocks are normalized such that they lead to the same initial response of output as the transitory VAR innovation. Figures 2 to 5 depict posterior mode responses for the model specifications as well as 75 percent highest posterior density bands based on the vector autoregression.

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, for instance, by Blanchard and Quah (1989) and Cochrane (1994). 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 generates an important trend-reverting component in output."

We examine the impulse response of the economy to a decrease in B_t, that is, an increase in the marginal rate of substitution between goods and leisure. The two right panels of Figure 2 depict the responses of the standard RBC model. The impulse responses show a typical business-cycle-like expansion: output, and hours increase. However, there is no trend-reverting output dynamics as indicated by the posterior distribution of the IRFs. The output response is monotonically decreasing after the initial increase. Figures 3 to 5 show the responses of the model economies with learning by doing. Output exhibits a clear hump-shaped response. However, the nature of the response is very sensitive to the parametrization of the skill accumulation equation. For large ϕ (Priors 1 and 2), the decay of output is too slow. For $\phi = 0.4937$ and $\mu = 0.2585$ it is much faster and the model impulse response function resembles the posterior predictions from the VAR more closely. Moreover, the hours response also exhibits a slight hump.

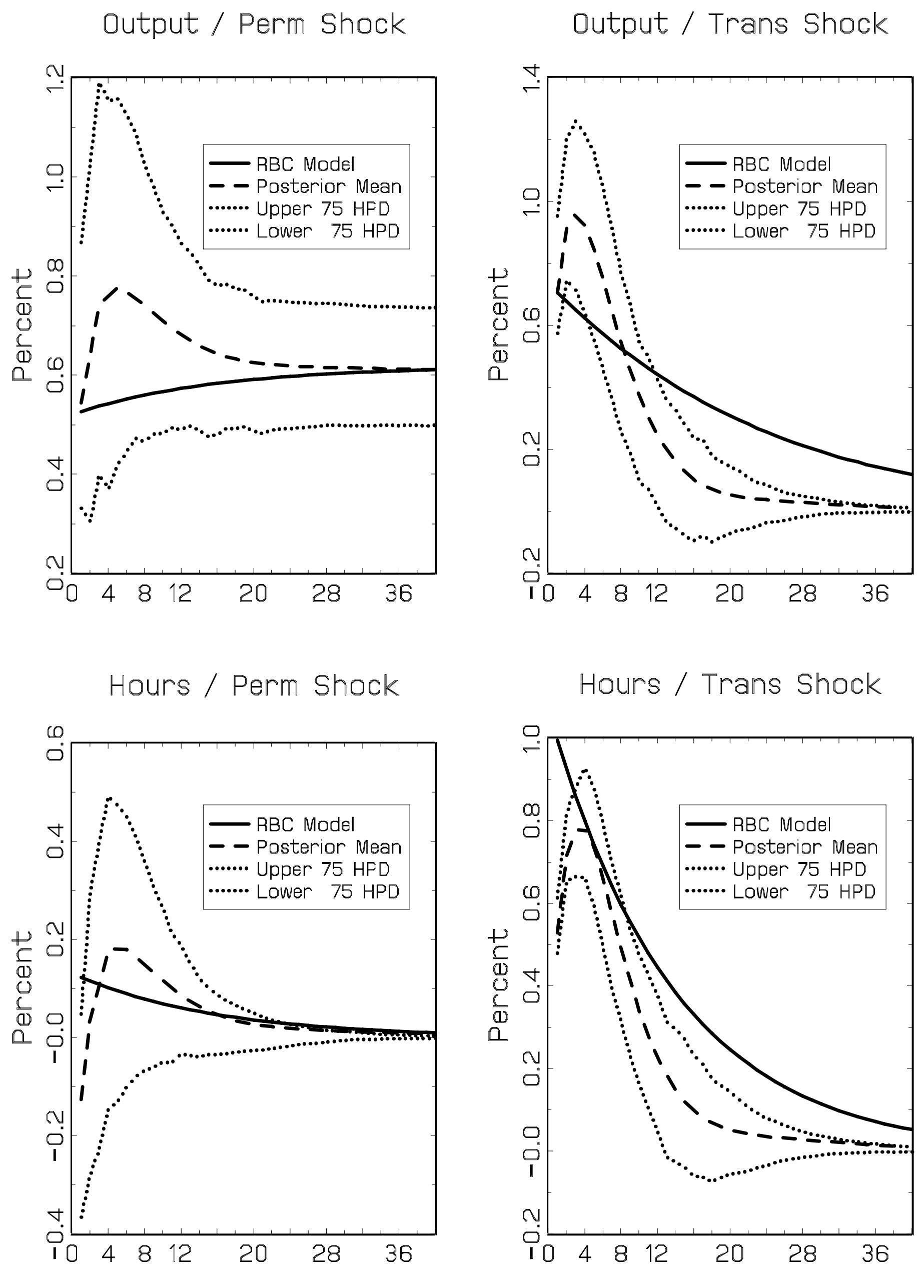

Figure 2: Impulse Response Functions: Posterior Distribution and Normalized Prediction of RBC Model.

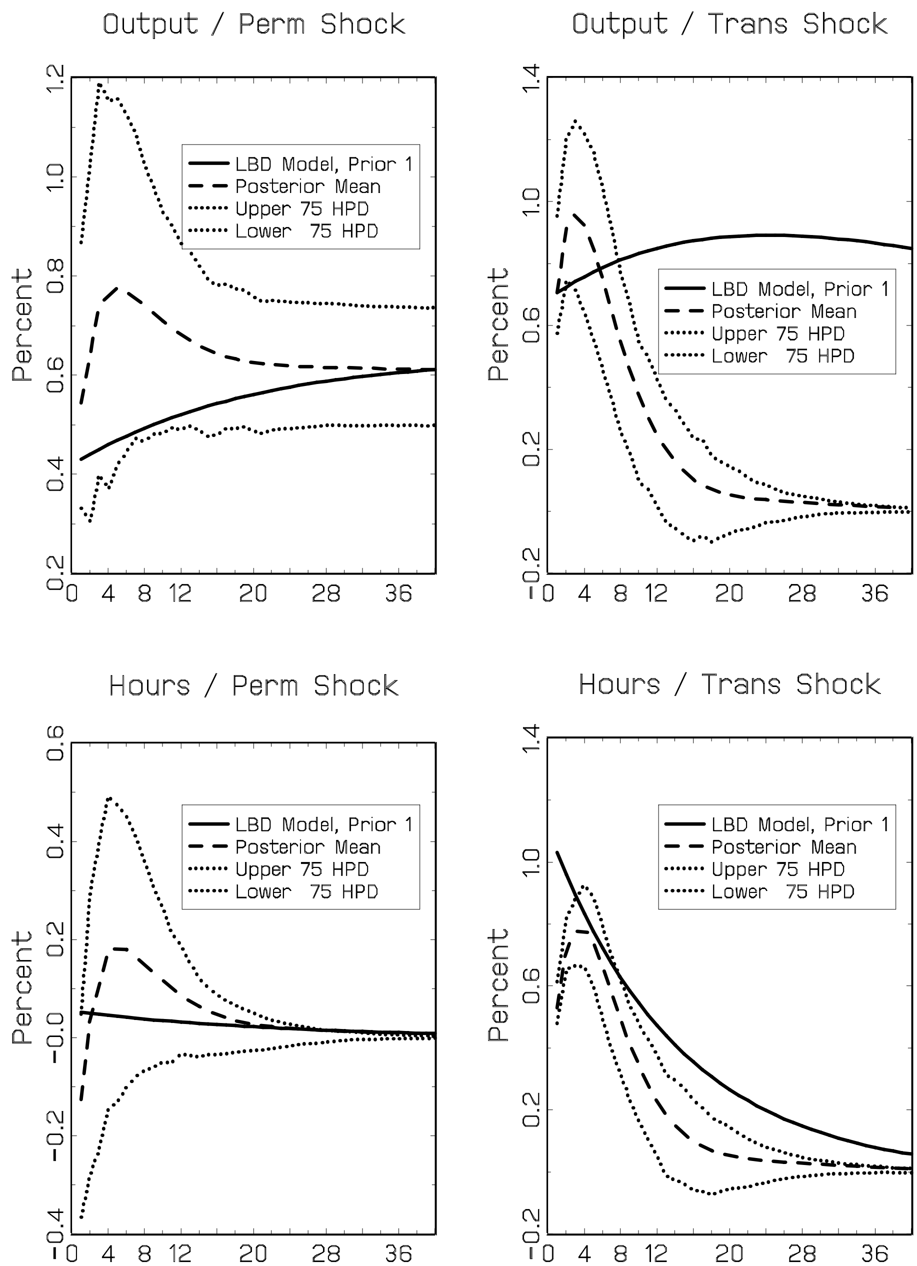

Figure 3: Impulse Response Functions: Posterior Distribution and Normalized Prediction of LBD Model with Prior 1.

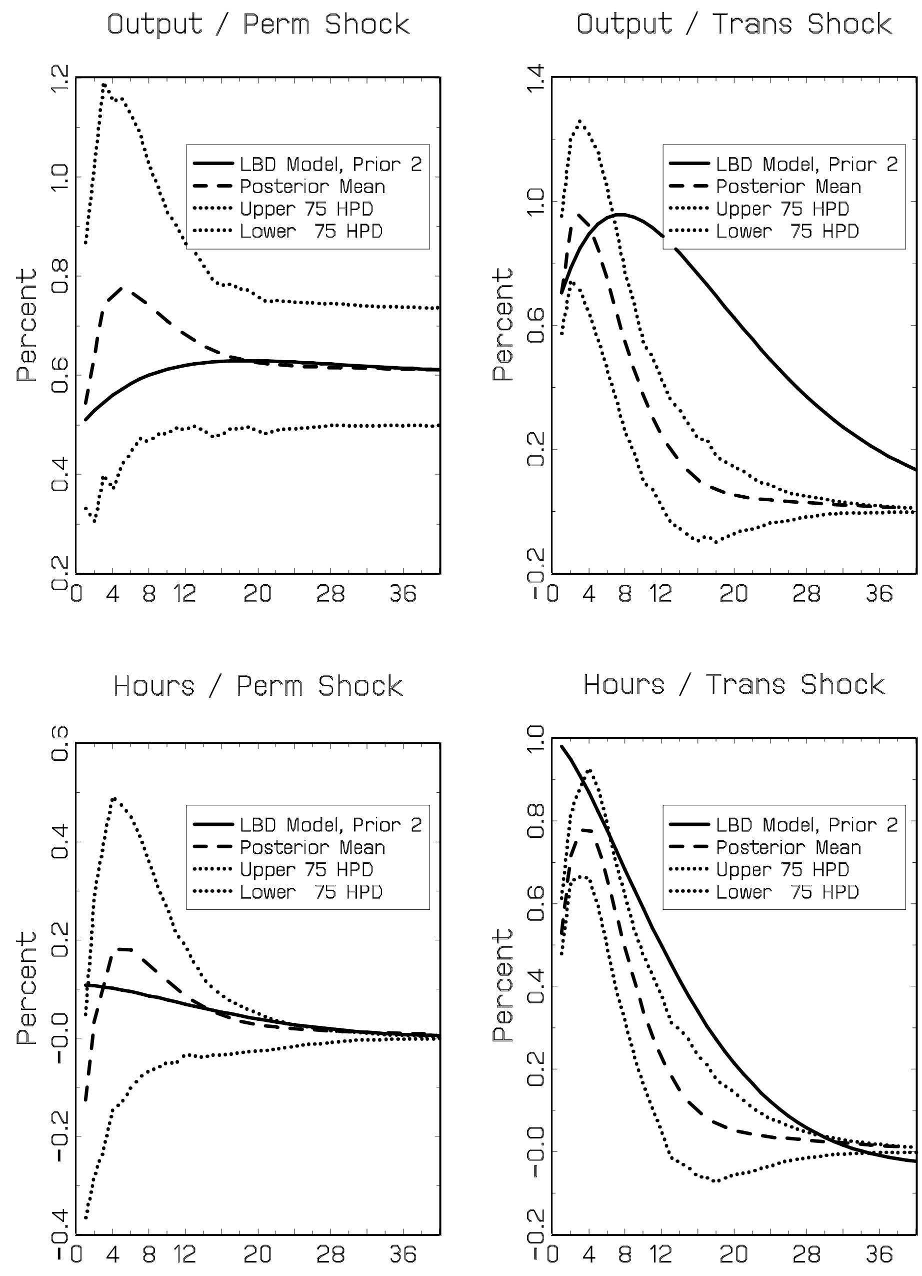

Figure 4: Impulse Response Functions: Posterior Distribution and Normalized Prediction of LBD Model with Prior 2.

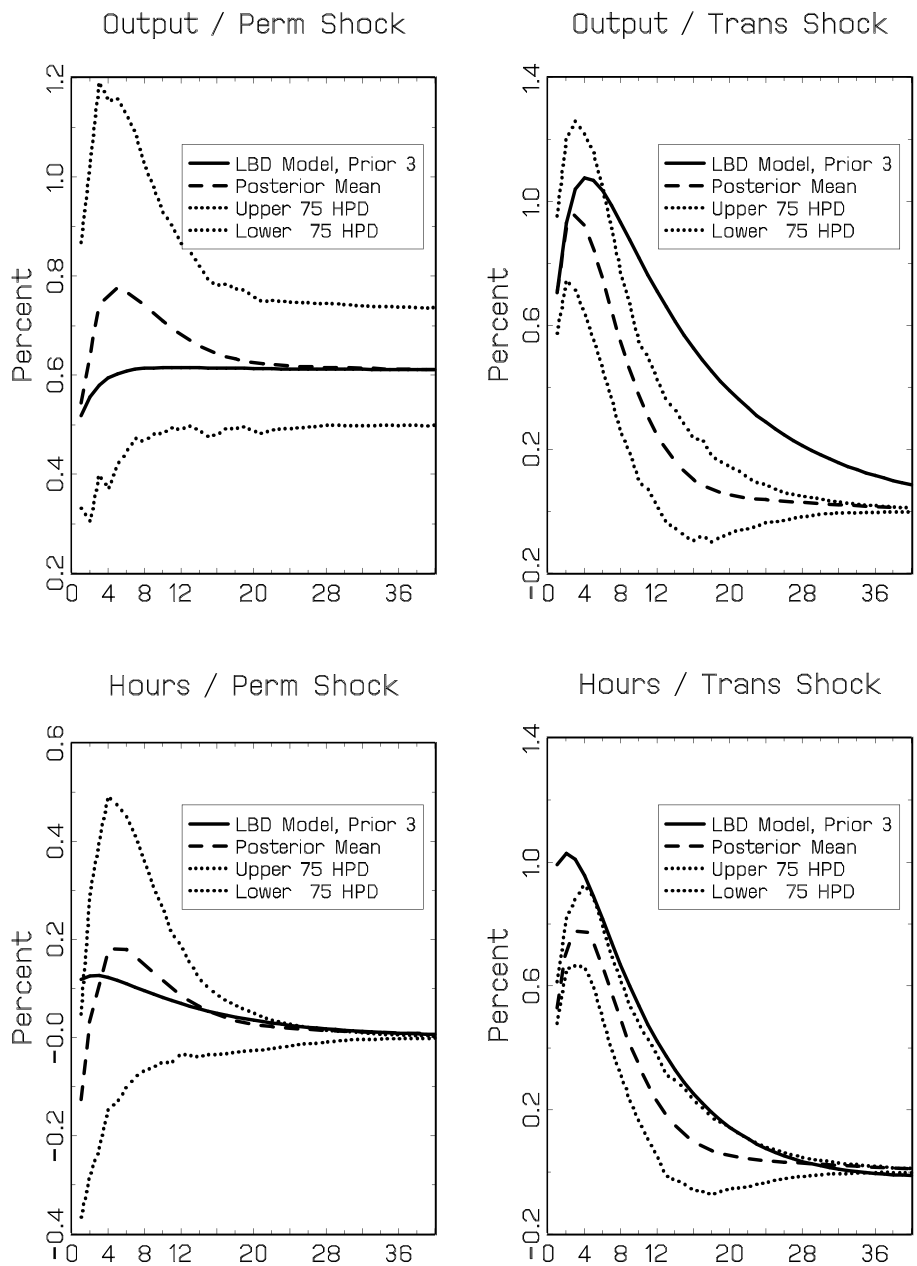

Figure 5: Impulse Response Functions: Posterior Distribution and Normalized Prediction of LBD Model with Prior 3.

4.3.5 Loss Function Estimation

In this section we ask the following question: is there a reasonable parametrization of the learning-by-doing model that produces impulse response functions of a similar shape as the posterior mean responses obtained from the VAR? Define the vector of population characteristics $\varphi = [\varphi^{(1)}, \varphi^{(2)}]'$ where $\varphi^{(1)}$ corresponds to the four impulse response functions for periods 1 to 40, and $\varphi^{(2)}$ to the structural parameters of the learning-by-doing model. We will use the modified quadratic loss function

$$L(\varphi, \hat{\varphi}) = (\varphi^{(1)} - \hat{\varphi}^{(1)})' \kappa W (\varphi^{(1)} - \hat{\varphi}^{(1)}) - \ln p(\varphi^{(2)} | \mathcal{M}_2) \tag{24}$$

to obtain parameter estimates for the LBD model. We chose the weight matrix W to be the identity matrix and $\kappa = 100$. The second term penalizes strong deviations from an *a priori* plausible parametrization of the structural model. If κ is large than the loss is dominated by the accuracy of the impulse response predictions. If $\kappa = 0$, the loss is minimized by the prior mode parameters.

Table 6 summarizes the results. At first glance, the parameter estimates look reasonable. The persistence of the skill accumulation process is somewhat higher than under the preferred specification $\mathcal{M}_{2(3)}$. The in-sample fit of the LBD model conditional on the loss function estimates is slightly worse than the fit of the other specifications. The log-likelihood decreases by about 20 points, relative to $\mathcal{M}_{2(3)}$. The L_{χ^2} losses are also slightly higher than for $\mathcal{M}_{2(3)}$. Figure 6 depicts the attained fit for the impulse responses. Both output responses exhibit hump shaped behavior. The response of output to the transitory shock decays much more rapidly than under $\mathcal{M}_{2(3)}$ and closely resembles the posterior mean response.

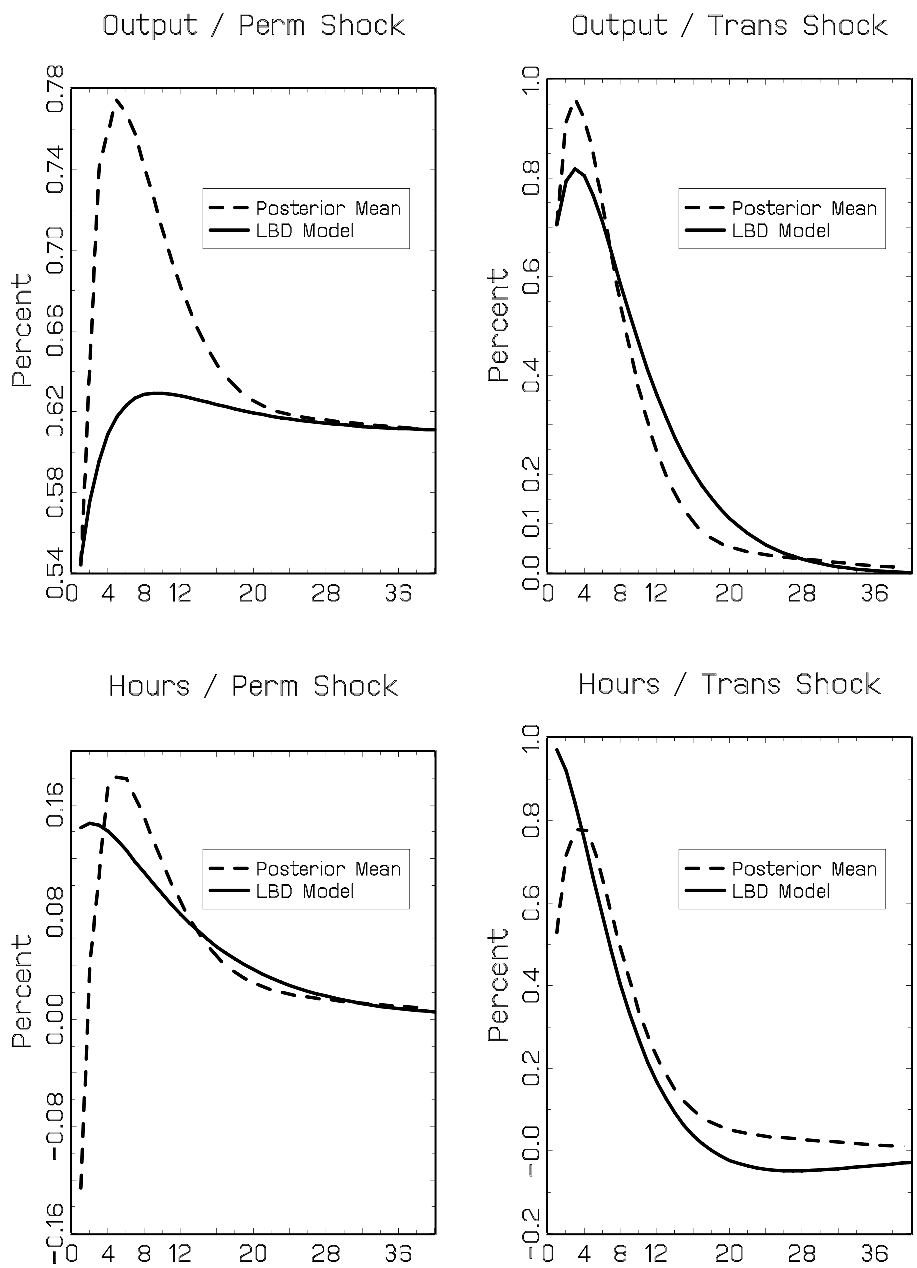

Figure 6: Impulse Response Functions: Posterior Distribution and Normalized Prediction of LBD Model with Loss Function Estimates.

Parameters		Model Features	
α	0.7261	Likelihood	1014.40
β	0.9900	In-sample RMSE Δy_t	0.00949
γ	0.0048	In-sample RMSE L_t	0.00855
δ	0.0250		
L_m	0.3177	$\sigma(\Delta y_t)$	0.0120
ρ	0.8759	$\sigma(L_t)/\sigma(\Delta y_t)$	1.8637
μ	0.1475	$corr(\Delta y_t, L_t)$	0.3157
ϕ	0.6097	L_{χ^2}-loss: $corr(\Delta y_t, \Delta y_{t-h})$, Lags 1-4	0.8938
ψ	3.1554	L_{χ^2}-loss: $corr(\Delta y_t, \Delta y_{t-h})$, Lags 1-8	0.7370
σ_a	0.0107	L_{χ^2}-loss: $corr(L_t, L_{t-h})$, Lags 1-4	1.0000
σ_b	0.0091	L_{χ^2}-loss: $corr(L_t, L_{t-h})$, Lags 1-8	0.9966

Table 6: IRF Loss Function Estimates for the LBD Model. The estimates for ψ, σ_a, and σ_b are obtained by maximizing the likelihood function conditional on the loss function estimates of the other parameters.

5 Conclusion

Despite its popularity and wide application, standard real business cycle models lack a persistent internal propagation mechanism. Aggregate output fluctuations simply trace out the movements of the exogenously given productivity process. To generate persistence, we augment the RBC model with a learning by doing mechanism, where current labor supply affects workers' future labor productivity. This is consistent with microeconomic evidence.

Based on the PSID data set we construct micro-level estimates for the parameters of the LBD mechanism. These estimates together with empirical evidence from related studies are combined with time series data on GDP growth and employment to perform a Bayesian analysis of the representative agent model. This analysis shows that the LBD model fits aggregate data much better than the standard RBC model. Posterior odds favor the learning by doing specification. Moreover, we formally showed that the LBD model more closely mimics the autocorrelation and impulse response patterns that we found in a bivariate VAR analysis. We view the learning by doing mechanism as a feature that can easily be build into more complicated dynamic stochastic general equilibrium models to improve their empirical performance. We incorporated the LBD feature into a home production model and were able to achieve improvements similar to the ones reported in this paper.

References

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

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

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

[4] Chang, Y. and N. Kwark (1999) "Decomposition of Hours based on Extensive and Intensive Margins of Labor," *Manuscript*, University of Pennsylvania.

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

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

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

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

[9] Den Haan, W., G. Ramey, and J. Watson (1997) "Job Destruction and Propagation of Shocks," *Manuscript*, University of California, San Diego.

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

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

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

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

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

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

[16] Rotemberg, J., and M. Woodford (1996) "Real Business Cycle Models and the Forecastable Movements in Output, Hours, and Consumption," *American Economic Review,* 86, 71-89.

[17] Ruhm, C. (1991) "Are Workers Permanently Scarred by Job Displacement" *American Economic Review,* 81, 319-324.

[18] Schorfheide, F. (1999) "A Unified Econometric Framework for the Evaluation of DSGE Models". *IER Working Paper*, University of Pennsylvania *http://www.econ.upenn.edu/~schorf/research.htm.*

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

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

A Data Sets

A.1 Micro Level Data

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 also use workers who were employed in non-agricultural sectors and not self-employed. Since we use wages and hours of four and five periods in the past as instruments, workers have to be employed at least at time t, $t-1$, $t-2$, $t-4$ and $t-5$ to make one observation. This gives us 24004 observations. Descriptive statistics for the sample used in the estimation are reported in Table 7.

A.2 Aggregate Data

The following time series are extracted from DRI: gross domestic product (GDPQ), employed civilian labor force (LHEM), civilian noninstitutional population, 20 years and older (PM20 and PF20). We defined $POP = PF20 + PM20$. From the BLS we obtained the series: average weekly hours, private non-agricultural establishments (EEU00500005). Prior to 1963 the BLS series in annual. We used these annual averages as monthly observations without further modification. Our monthly measure of hours worked is

$$BLShours_c = EEU00500005 * LHEM \ / \ POP$$

We convert to quarterly frequency by simple averaging. Per capita output is

$$GDPQ_c = GDPQ \ / \ POP$$

Output growth is defined as $\ln GDPQ_c(t) - \ln GDPQ_c(t-1)$.

B Estimation and Computation

B.1 Conversion of Estimates from Annual to Quarterly

Let x_t denote the skill level at quarterly frequency:

$$x_t = \phi x_{t-1} + \mu h_t \tag{25}$$

Variable	Mean	Std.D.	No. of Obs.
Real Wage (in \$1983 dollars)	12.39	7.27	24004
Annual Hours of Work	2113.70	508.50	24004
Age	41.10	10.14	24004
Years of Schooling	13.27	2.49	23865
Gender (male=1)	0.67	0.47	24004

Table 7: Descriptive Statistics for PSID Sub-Sample.

Suppose that $h_t = h$, where h is the steady state level of weekly hours at work. Then

$$x_t - \mu h/(1-\phi) = \phi^t[x_0 - \mu h/(1-\phi)] \tag{26}$$

Denote annual average hourly earnings by

$$\tilde{x}_t = (x_{4t} + x_{4(t-1)} + x_{4(t-2)} + x_{4(t-3)})/4 \tag{27}$$

For fixed $h_t = h$ we obtain

$$\begin{aligned}
\tilde{x}_2 &= (x_8 + x_7 + x_6 + x_5)/4 \\
&= \frac{\mu h}{1-\phi} + \frac{1}{4}\phi^4[\phi^4 + \phi^3 + \phi^2 + \phi]\left(x_0 - \frac{\mu h}{1-\phi}\right) \\
&= \phi^4 \tilde{x}_1 + \frac{1}{4}\frac{1-\phi^4}{1-\phi}\mu 4h
\end{aligned} \tag{28}$$

where $4h$ are the annualized hours worked. This implies the following relationship between coefficients for quarterly data and coefficients for annual data

$$\phi = \tilde{\phi}^{1/4}, \quad \mu = 4\frac{1-\phi}{1-\phi^4}\tilde{\mu} \tag{29}$$

Point estimates and standard errors for quarterly frequency are computed via linearization of Equation (29) around $\hat{\tilde{\phi}}$ and $\hat{\tilde{\mu}}$.

B.2 Vector Autoregression

A fourth order vector autoregression serves as reference model:

$$y_t = C_0 + C_{tr}t + \sum_{h=1}^{4} C_h y_{t-h} + u_t \quad u_t \sim \mathcal{N}(0, \Sigma) \tag{30}$$

where y_t denotes a vector of GDP growth and hours worked. The structural models imply that both series are stationary. However, we include a deterministic time trend since the empirical measure of hours worked does exhibit some trending behavior. Define the $1 \times k$ vector $x_t = [1, t, y'_{t-1}, \ldots, y'_{t-4}]$ and the matrix of regression coefficients $C = [C_0|C_{tr}|\ldots|C_4]'$. 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

$$y_{it} = C_{i,0} + C_{i,tr}t + C_{i,1}y_{t-1} + \cdots + C_{i,4}y_{t-4} + u_t, \quad i = 1, 2 \tag{31}$$

Since y_{1t} is differenced output and the theory implies that hours are stationary we chose the prior mean to be zero for all coefficients. The prior variance for C_{i0} and C_{iT} is large 10^6, that is, the prior of these coefficient is flat. The variance of C_{ijl}, $l = 1, \ldots, p$ is given by

$$var(C_{ijl}) = \begin{cases} (\zeta/l)^2 & \text{if} \quad i = j \\ (\zeta\hat{\sigma}_i/l\hat{\sigma}_j)^2 & \text{if} \quad i \neq j \end{cases} \tag{32}$$

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

$$vec(C) \sim \mathcal{N}\Big(vec(\bar{C}), V_c(\zeta)\Big) \tag{33}$$

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 chose $\zeta_1 = 0.0001$, $\zeta_J = 10$, $J = 20$, and $\ln\zeta_j$ equally spaced in the interval $[\ln\zeta_1, \ln\zeta_J]$.

B.2.2 Gibbs Sampling

The Gibbs sampler is used to obtain draws $(C^{(s)}, \Sigma^{(s)}, \zeta^{(s)})$, $s = 1, \ldots, n_{sim}$ from the posterior distribution $p(C, \Sigma, \zeta|Y, \mathcal{M}_*)$. of the VAR parameters. Define

$$\hat{C} = (X'X)^{-1}X'Y \tag{34}$$

$$\hat{\Sigma} = (Y - X\hat{C})'(Y - X\hat{C})/T \tag{35}$$

The following conditional posterior distributions are used for the Gibbs sampling algorithm.

1. The conditional probability $I\!P[\{\zeta = \zeta(j)|Y, C, \Sigma\}]$ is given by

$$
\begin{aligned}
& I\!P[\{\zeta = \zeta(j)|Y, C, \Sigma\}] \\
& \quad = \frac{|V_c(\zeta(j))|^{-1/2} \exp\left\{-\frac{1}{2}(c-\bar{c})' V_c^{-1}(\zeta(j))(c-\bar{c})\right\}}{\sum_{j=1}^{J} |V_c(\zeta(j))|^{-1/2} \exp\left\{-\frac{1}{2}(c-\bar{c})' V_c^{-1}(\zeta(j))(c-\bar{c})\right\}}
\end{aligned}
$$

2. The conditional density $p(C|Y, \Sigma, \zeta)$ has the shape of a multivariate normal density with mean $\tilde{c}(\zeta)$ and covariance matrix $\tilde{V}_c(\zeta)$.

$$
\begin{aligned}
\tilde{c}(\zeta) & \equiv [V_c^{-1}(\zeta) + (\Sigma^{-1} \otimes X'X)]^{-1}[V_c^{-1}(\zeta)\bar{c} + (\Sigma^{-1} \otimes X'X)\hat{c}] \\
\tilde{V}_c^{-1}(\zeta) & \equiv V_c^{-1}(\zeta) + (\Sigma^{-1} \otimes X'X)
\end{aligned}
$$

3. The conditional density $p(\Sigma|Y, C, \zeta)$ is of the Inverted Wishart type with T degrees of freedom and parameter H where

$$H = T\hat{\Sigma} + (C - \hat{C})' X' X (C - \hat{C})$$

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 $p(\varphi|data)$.

B.2.3 Marginal Data Density

We compute marginal data densities conditional on a training sample of 18 observations. The conditional data density is proper and can be used to obtain posterior model probabilities. The marginal data density can be expressed as

$$p(Y_T|Y_0, \mathcal{M}_*) = \prod_{t=1}^{T} \int p(y_t|Y_{t-1}, Y_0, C, \Sigma, \zeta, \mathcal{M}_*) p(C, \Sigma, \zeta|Y_{t-1}, \mathcal{M}_*) d(C, \Sigma, \zeta) \quad (36)$$

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 20000 draws to approximate the integral.

B.3 Structural Models

Conditional on the actual data Y_T, Y_0, a set of parameter values θ_i, the Kalman Filter is used to evaluate the posterior density up to a constant

$$p(\theta_i|Y_T, Y_0, \mathcal{M}_i) \propto p(Y_T|Y_0, \theta_i, \mathcal{M}_i) p(\theta_i|\mathcal{M}_i) \quad (37)$$

A numerical optimization routine is used to compute the mode $\tilde{\theta}_i$ of the posterior density. Let $\tilde{\Sigma}_i$ be

$$\tilde{\Sigma}_i = \left[-\frac{\partial^2}{\partial\theta_i \partial\theta_i'} \ln p(Y_T|Y_0, \theta_i, \mathcal{M}_i) p(\theta_i|\mathcal{M}_i) \right]_{\theta_i = \tilde{\theta}_i} \tag{38}$$

the Hessian of the log posterior density evaluated at the mode. The reported standard errors correspond to the square roots of the diagonal elements of $\tilde{\Sigma}_i$. The marginal data densities are calculated by Laplace approximation:

$$\tilde{p}(Y_T|Y_0, \mathcal{M}_i) = (2\pi)^{d/2} |\tilde{\Sigma}_i|^{1/2} p(Y_T|Y_0, \tilde{\theta}_i, \mathcal{M}_i) p(\tilde{\theta}_i|Y_0, \mathcal{M}_i) \tag{39}$$

which is based on a log-quadratic expansion around the posterior mode.