Costly Vs. By-product LBD model: A Bayesian Evaluation

preliminary draft

Keqiang Hou

Department of Economics
McMaster University
Hamilton, Ontario
E-mail *houk2@mcmaster.ca*

Alok Johri

Department of Economics
McMaster University
Hamilton, Ontario
E-mail *johria@mcmaster.ca*

December 8, 2008

Contents

Abstract

A key aspect of the typical formulation of DGE models with learning-by-doing (LBD) is that firms do not need to incur resources to add to their stock of organizational capital. We construct a tractable model with an alternate approach to learning-by-doing in which firms must spend considerable resources to learn how best to combine their inputs and raise future productivity. We refer to this view as costly LBD as opposed to by-product LBD. We then take the two models to the aggregate U.S. data using a Bayesian method and compare their quantitative implications. We find that both models, give fairly similar results, both for posterior estimates, as well as the implied posterior moments and forecast error variance. While either form of LBD is preferred to the model without it, the marginal data density is in favor of byproduct model over the costly model by a slim margin.

1 Introduction

The idea that economic activity involves learning-by-doing (lbd) has played an important role in a number of economic literatures including growth theory and business cycle analysis as well as industrial organization and labour economics. While learning-by-doing is often associated with workers and modeled as the accumulation of human capital, a number of economists have argued that firms are also store-houses of knowledge. Atkeson and Kehoe (2005) note "At least as far back as Marshall (1930, bk.iv, chap. 13.I), economists have argued that organizations store and accumulate knowledge that affects their technology of production. This accumulated knowledge is a type of unmeasured capital distinct from the concepts of physical or human capital in the standard growth model." Similarly Lev and Radharkrishnan (2003) write, "Organization capital is thus an agglomeration of technologies—business practices, processes and designs, including incentive and compensation systems—that enable some firms to consistently extract out of a given level of resources a higher level of product and at lower cost than other firms."[1]

A key aspect of the typical formulation of learning-by-doing is that knowledge accumulation occurs as a by-product of production which in turn leads to productivity increases. This formulation (whether it involves external effects or not) draws on early work by Arrow (1962) and Rosen (1972) as well as a large empirical literature dating back roughly a hundred years. That literature documents the pervasive presence of learning effects in virtually every area of the economy. Recent studies include Bahk and Gort (1993), Irwin and Klenow (1994), Jarmin (1994), Benkard (2000), Thornton and Thompson (2001), Chang, Gomes and Schorfheide (2002) and Cooper and Johri (2002).

An alternative approach to lbd acknowledges that firms in fact spend

[1]There are at least two ways to think about what constitutes organizational capital. Some, like Rosen (1972), think of it as a firm specific capital good while others focus on specific knowledge embodied in the matches between workers and tasks within the firm. While these differences are important, especially when trying to measure the payments associated with various inputs, they are not crucial to the issues at hand. As a result we do not distinguish between the two.

considerable resources to learn how best to combine their inputs and raise future productivity. We refer to this alternative view as costly lbd as opposed to by-product lbd. In this paper we write down a tractable model of costly lbd and provide aggregate estimates of both types of models using a Bayesian likelihood approach. Both models involve firms that operate in a monopolistically competitive environment and have access to a technology which allows the accumulation of production related knowledge which we refer to as organizational capital. The two models differ only in that firms in one model will have to choose how to allocate their resources between production and knowledge accumulation in a dynamically optimal way.

The paper estimates these dynamic stochastic general equilibrium (DSGE) models using quarterly data on total hours and aggregate output from the U.S. between 1954:II and 1997:IV. To confront the models with the data, we make use of Bayesian methods to combine prior judgments together with information contained in the historical aggregate data. The main results of this paper are as follows. First, introducing learning-by-doing significantly improves the overall likelihood-based fit of the model relative to the model without any form of lbd. Second, lbd serves as an important propagation mechanism for shocks. Any shock that leads to the accumulation of organizational capital implies that labor and capital will be more productive in the future than they would be in the absence of the lbd mechanism. Hence, output and hours display considerably more inertia in response to shocks than the standard model and persistence of output is significantly increased. Third, we find that the costly and byproduct models have very similar qualitative implications for aggregate variables to exogenous shocks. Both models drive a wedge between labor productivity and wages but through different channels. The by-product LBD model delivers an endogenous time varying price-cost markup, while the costly LBD model drives a time varying wedge between the wage and the marginal product of labour. Finally, the posterior estimates of structural parameters are similar in the two learning models. However, the marginal data density is marginally in favor of the by-product LBD model over the costly LBD model.

The rest of the paper is organized as follows. Section 2 lays out the basic structure of our model economy with different specifications of learning-by-doing. Section 3 discusses the econometric methodology and the data and then presents the empirical results. Section 4 concludes.

2 The Model Economy

In this section, we specify three related models as following. First, if learning dynamics is abstracted from producers and organizational capital will be chosen to be a constant, then the model reduces to the standard RBC model of monopolistic competition. For convenience we refer to this specification as the *benchmark model*. Second, when organizational capital occurs as the byproduct of output production, the model becomes observationally equivalent to the model in Clark and Johri (2008). For convenience we refer to this as the *byproduct model*. Finally, when assuming firms need to pay considerable economic price to engage in organizational capital production, the model proposes the alternative view of learning dynamics as opposed to byproduct model. For convenience we refer to this model as the *costly model*. The three competing models disagree with each other regarding the key aspect of the formulation of learning-by-doing at the firm level but they share the common assumptions upon the preference and final goods sector. Thus we first describes the set of equations that is common to all three models and then discuss the intermediate-good sectors, which are different for each model.

2.1 Households

The economy is populated by a large number of identical households. The preferences of the representative household are defined over consumption of final goods and leisure. The representative household maximizes the expected discounted utility over an infinite life horizon by choosing C_t, labor supply N_t, and investment in physical capital I_t, taking as given the real wage w_t and the real capital return r_t^k. Households supply labor services and rent physical capital to the intermediate goods produc-

ers. Households also owns these intermediate firms and receive real dividends payment $\pi_t(i)$ from each intermediate firm $i \in [0,1]$. The representative household maximizes her intertemporal utility function given by:

$$\max \sum_{t=0}^{\infty} \beta^t U(C_t, N_t, \mathcal{B}_t), \tag{1}$$

where β is the discount factor. The utility function at period t depends positively on the contemporaneous consumption of final goods, C_t and labor supply, N_t. (1) contains a preference shock $\mathcal{B}_t$, which represents a taste shock to labor supply and follows a first-order autoregressive process with an iid error term:

$$\ln \mathcal{B}_t = \rho_b \ln \mathcal{B}_{t-1} + \epsilon_{bt} \tag{2}$$

The representative household maximizes her objective function subject to her sequence of budget constraints given by

$$\begin{aligned} C_t + I_t + \mathrm{E_t} Q_{t,t+1} B_{t+1} = \\ w_t N_t + r_t^k K_t + B_t + \int_0^1 \pi_t(i) di, \quad t = 0, 1, \ldots, \infty \end{aligned} \tag{3}$$

where B_{t+1} is one-period securities with price $Q_{t,t+1}$ and $Q_{t,t+1}$ denotes the period-t price of a claim to one unit of final goods in period $t+1$. The Household holds her financial wealth in the form of bonds B_{t+1}. Note that the borrowing constraints $B_{t+1} \geq \bar{B}$ for some large negative number $\bar{B}$. The right-hand side of the budget constraint represents the sources of wealth: labor income $w_t N_t$[2]; the return on the real capital stock, $r_t^k K_t$, the payoff of contingent claims acquired in the previous period B_t and the dividends derived from the imperfect competitive intermediate firms.

[2]The households supplies $N_t(i)$ units of labor to each intermediate good producer and earns factor payment in total $w_t N_t$ in period t, where the total hours worked N_t must satisfy

$$N_t = \int_0^1 N_t(i) di$$

The left-hand side shows the uses of wealth: consumption spending, investment in physical capital and purchases of interest bearing assets. In addition, Investment augments the physical capital stock over time according to

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

where $\delta \in (0,1)$ is a constant depreciation rate for physical capital.

Given initial values, the household chooses $\{C_t, N_t, I_t, K_{t+1}, B_{t+1}\}, t = 0,1,2,\ldots$, to maximize the objective function (1) subject to the budget constraint (3) and the capital accumulation equation (4). The first-order conditions associated with this problem are:

$$w_t = \mathcal{B}_t \frac{U_{n,t}}{U_{c,t}} \tag{5}$$

$$1 = \beta E_t \left[\frac{U_{c,t+1}}{U_{c,t}} \left(r^k_{t+1} + 1 - \delta \right) \right] \tag{6}$$

$$\frac{1}{R_t} = \beta E_t \left(\frac{U_{c,t+1}}{U_{c,t}} \right) \tag{7}$$

where R_t is the gross rate of return on bonds ($R_t = 1 + r^b_t = E_t \frac{1}{Q_{t,t+1}}$) and $U_{c,t}$ and $U_{n,t}$ are, respectively, the marginal utility of consumption and marginal utility of leisure. Equation (5) gives us the intra-temporal optimality condition, which equate the marginal rate of substitution between consumption and labor to the real wage. The preference shock $\mathcal{B}_t$ here play a role of shifting the marginal rate of substitution. Equation (6) is the standard Euler equation for the accumulation of physical capital which states that, at the optimal, the utility cost of sacrificing one unit of consumption must be equal to discounted utility benefit of this unit consumption tomorrow, while equation (7) gives the Euler equation for inter-temporal consumption.

2.2 Final Good Producers

There are large number of final good producers who behave competitively and use $y_t(i)$ units of a continuum of intermediate good $i \in [0,1]$, to

produce Y_t units of the final good. Assuming that all intermediate goods are imperfect substitutes with a constant elasticity of substitution, $\frac{\eta}{\eta-1}$, the corresponding Dixit-Stiglitz aggregator can be defined as:

$$Y_t = \left[\int y_t(i)^{\eta} \, di \right]^{\frac{1}{\eta}}, \qquad \eta > 1 \tag{8}$$

Given the relative price vector, the final-good producer chooses the quantity of intermediate good $Y_t(i)$ that maximizes its profits,

$$\max_{y_t(i)} \; Y_t - \int v_t(i) y_t(i) \, di \tag{9}$$

subject to the constraint imposed by (8). Note that $v_t(i)$ is the relative price charged by the ith intermediate goods producers. The first order conditions for this problem give us the input demand functions:

$$Y(i, s^t) = v_t(i)^{-\frac{\eta}{\eta-1}} Y_t \tag{10}$$

where $\frac{\eta}{\eta-1}$ measures the price elasticity of demand for intermediate good i.

2.3 Intermediate Good Producers

The above sections describe the set of equations that is common to the models of interest, while this section discusses the intermediate goods producer problems, which are different for each model.

2.3.1 Costly Model Specification

The economy produces a continuum of intermediate goods indexed by $i \in [0, 1]$. Each intermediate good i is produced by firm i using the following technology:

$$y_t(i) = (A_t u_t^n(i) N_t(i))^{\alpha} \left(u_t^k K_t(i) \right)^{1-\alpha} Z_t(i)^{\varepsilon} \tag{11}$$

where organizational capital, $Z_t(i)$, is another factor input combined with labor $N_t(i)$ and physical capital $K_t(i)$ to produce intermediate goods $y_t(i)$

and the variable, $u_t^n(i)$ denotes the faction of labor which firm i choose to use in output production activities. Firms will use the rest of labor to engage in organizational capital accumulation. A_t is the productivity shock that is common to all intermediate good producers. The technology shock, A_t, is assumed to follow a random walk with drift process:

$$\ln A_t = \gamma_a + \ln A_{t-1} + \epsilon_{at} \tag{12}$$

where ϵ_{At} is *iid* shocks. Intermediate good producers use up certain amount of physical capital, labor and the stock of organizational capital to internally accumulate the stock of organizational capital for the next period. We specify the accumulation equation as

$$Z_{t+1}(i) = \left[\left(A_t(1-u_{i,t}^n) N_t(i) \right)^{\alpha_1} \left((1-u_{i,t}^k) K_t(i) \right)^{1-\alpha_1} \right]^{(1-\gamma)} Z_t(i)^{\gamma} \tag{13}$$

where α_1 represents the elasticity of hours worked in current period with respect to organizational capital in the next period and $\gamma \in (0,1)$ indicates that organizational capital accumulation is persistent but not permanent. The stock of organizational capital decays at the rate γ, which is consistent with the empirical evidence supporting the hypothesis of depreciation of organizational capital.

In contrast to previous learning-by-doing literature where organizational capital is a byproduct of production which in turn leads to productivity increases, we require firms to pay considerable economic price to learn how best combine their inputs and raise future productivity in costly specification. In the market for intermediate-good, each differentiated intermediate good i is produced by a single firm i. Thus a producer i solve his maximization problem by choosing contingency plans for $\{v_t(i), u_t^n(i), u_t^k(i), N_t(i), K_t(i), Z_{t+1}(i)\}_{t=0}^{\infty}$ that maximize his present value of real dividends:

$$\mathcal{X}_t(i) = \max \sum_{t=0}^{\infty} D_t \left(v_t(i) y_t(i) - w_t N_t(i) - r_t^k K_t(i) \right)$$

subject to the input demand function (10), output production technology (11), organizational capital accumulation function (13) and an appropriate

transversality condition on the stock of organizational capital. Since firms are owned by the household, D_t is the appropriate endogenous discount rate for the firms. The first order conditions are given by:

$$w_t - \alpha\lambda_t^y \frac{y_t(i)}{N_t(i)} - \alpha_1\lambda_t^Z \frac{Z_{t+1}(i)}{N_t(i)} = 0 \quad (14)$$

$$r_t^k - (1-\alpha)\lambda_t^y \frac{y_t(i)}{K_t(i)} - (1-\gamma-\alpha_1)\lambda_t^Z \frac{Z_{t+1}(i)}{K_t(i)} = 0 \quad (15)$$

$$\alpha\lambda_t^y \frac{y_t(i)}{u_t^n(i)} - \alpha_1\lambda_t^Z \frac{y_t(i)}{1-u_t^n(i)} = 0 \quad (16)$$

$$(1-\alpha)\lambda_t^y \frac{y_t(i)}{u_t^k(i)} - (1-\gamma-\alpha_1)\lambda_t^Z \frac{y_t(i)}{1-u_t^k(i)} = 0 \quad (17)$$

$$\lambda_t^Z - E_t\left[\frac{D_{t+1}}{D_t}\left(\varepsilon\lambda_{t+1}^y \frac{y_{t+1}(i)}{Z_{t+1}(i)} + \gamma\lambda_{t+1}^Z \frac{Z_{t+2}(i)}{Z_{t+1}(i)}\right)\right] = 0 \quad (18)$$

$$v_t(i) - \eta\lambda_t^y(i) = 0 \quad (19)$$

where λ_t^y and λ_t^Z are the Lagrangian multipliers associated with constraints (11) and (13), respectively. Equation (14) and (15) equate the marginal production of labor and physical capital, respectively, to their factor prices. Equation (16/17) states that firm i should choose the fraction of labor/capital in such a way that revenue of using an additional unit of labor/capital in output production can be exactly offset by the revenue of using this unit of labor/captial to produce organizational capital for the next period. Equation (18) shows that the cost producing an additional unit of organizational capital today must be equal to the discounted benefit of this additional organizational capital tomorrow. Equation (19) indicates that an intermediate-good producer chooses its relative price for its differentiated goods as a constant markup over the real marginal cost. Note that monopolistic firms maximize their profits by equating the marginal revenue to the marginal cost. Thus, λ_t^y can be interpreted as either marginal revenue or marginal cost at the optimum.

2.3.2 By-product Model Specification

In what follows, we offer a decentralization of Cooper and Johri(2002) where all learning occurs as a by-product of production at firms level. We

write the intermediate good production technology and the accumulation technology for the by-product organizational capital as follows[3]:

$$y_t(i) = (A_t N_t(i))^{\alpha} (K_t(i))^{1-\alpha} Z_t(i)^{\varepsilon}, \qquad \alpha, \varepsilon \in (0,1) \tag{20}$$

$$Z_{t+1}(i) = y_t(i)^{\eta} Z_t(i)^{\gamma}, \qquad \eta, \gamma \in (0,1) \tag{21}$$

A producer i solve its maximization problem in two stage. In the first stage, the producer chooses the cost minimizing quantities of labor and capital, for a given stock of organizational capital to solve the following static cost minimization problem:

$$\mathcal{C}_t(i) = \min_{N_t(i), K_t(i)} \left(w_t N_t(i) + r_t^k K_t(i) \right)$$

subject to (20). This cost minimization problem give us conditional factor demands that are function of factor price, output and the stock of organizational capital. After simple algebra, the minimized total cost, $\mathcal{C}_t(i)$, is given by:

$$\mathcal{C}_t(i) = \tilde{\Gamma} w_t^{\alpha} r_t^{1-\alpha} Z_t(i)^{-\varepsilon} y_t(i) \tag{22}$$

where $\tilde{\Gamma} = \left(1 + \frac{\alpha}{1-\alpha}\right) \left(\frac{\alpha}{1-\alpha}\right)^{-\alpha}$. The cost function (22) will be a non-increasing function of the (given) stock of organizational capital.

In the second stage, the intermediate goods producers will solve a dynamic problem that selects the contingency plans for $\{v_t(i), Z_{t+1}(i)\}_{t=0}^{\infty}$ that maximize the present value of real profits:

$$\max_{v_t(i), Z_{t+1}(i)} E_0 \sum_{t=0}^{\infty} D_t \left(v_t(i) y_t(i) - \mathcal{C}_t\{w_t, r_t^k, y_t(i), Z_t(i)\} \right)$$

subject to the demand function (10), the accumulation technology for by-product organizational capital (21) and an appropriate transversality condition on the stock of organizational capital. D_t is the appropriate endogenous discount rate for the firms.

[3]Cooper and Johri (2002) provide evidence on this specification

The solution to this maximization problem will satisfy the following first order conditions:

$$v_t(i)\frac{\partial y_t(i)}{\partial v_t(i)} + y_t(i) - mc_t(i)\frac{\partial y_t(i)}{\partial v_t(i)} + \lambda_t^Z(i)\frac{\partial Z_{t+1}(i)}{\partial y_t(i)}\frac{\partial y_t(i)}{\partial v_t(i)} = 0 \quad (23)$$

$$\lambda_t^Z(i) - E_t\left[\frac{D_t+1}{D_t}\left(\lambda_{t+1}^Z(i)\frac{\partial Z_{t+2}(i)}{\partial Z_{t+1}(i)} - \frac{\partial \mathcal{C}(i)}{\partial Z_{t+1}(i)}\right)\right] = 0 \quad (24)$$

where $mc_t(i)$ denotes the marginal cost of producing output $y_t(i)$. The term $\lambda_t^Z(i)$ denotes the Lagrange multiplier associated with the constraint (21) and represents the discounted value of an additional unit of organizational capital in terms of the real profits to the producer.

The first order condition (23) captures the nature of the dynamic tradeoff that arises when intermediate goods producers face a downward sloping demand curve. Under such circumstances, producers raise the output relative price by one unit causes demand for their product to fall, which leads to a decrease in their future productivity. With learning, intermediate good producers are allowed to do intratemporal and intertemporal arrangement regarding current profits and future productivity. The last term in (23), which is absent in the standard model of monopolistic competition, captures such tradeoff between maximizing current period profits and losing future productivity increase. Equation (24) simply implies that organizational capital will be accumulated up to the point where the value of an extra unit of organizational capital today is equal to the discounted value of this organizational capital next period.

2.4 Comparison between LBD models

The fundamental difference underlying the LBD models are their respective organizational capital accumulation mechanisms. Consequently, intermediate goods producers in the different models would derive different optimal decisions when facing a downward sloping demand curve. For producers in the byproduct model, they choose to vary the price-cost markups endogenously to capture the dynamic tradeoff between currently profit and future productivity. For producers in the costly model, however, they instead choose to vary wage markup endogenously when

such dynamic tradeoff arise. To see the key distinguishing feature between different learning dynamics explicitly, we rearrange the optimality conditions for costly and byproduct model.

- the Costly Model

We rewrite (13) as $N_t = \Psi(u_t^n, u_t^k, K_t, Z_t, Z_{t+1})$ and obtain the following optimality conditions for the costly model:

$$w_t = \mathcal{B}_t mrs_t \tag{25}$$

$$U_{c,t} = \beta E_t \left[U_{c,t+1} \left(r_t^k + 1 - \delta \right) \right] \tag{26}$$

$$v_{i,t} = \eta \lambda_t^{y_i} \tag{27}$$

$$w_t = \frac{1}{u_t^n} \cdot \lambda_t^{y_i} \alpha \frac{y_{i,t}}{N_{i,t}} \tag{28}$$

$$r_t^k = \frac{1}{u_t^k} \cdot \lambda_t^{y_i} (1-\alpha) \frac{y_{i,t}}{K_{i,t}} \tag{29}$$

$$\begin{aligned} &\left(\mathcal{B}_t mrs_t - u_t^n w_t\right) \Psi_t^{Z'} = \\ &\quad E_t \left\{ \frac{D_{t+1}}{D_t} \left[\lambda_{t+1}^{y_i} \varepsilon \frac{y_{i,t+1}}{Z_{i,t+1}} - \left(\mathcal{B}_{t+1} mrs_{t+1} - u_{t+1}^n w_{t+1}\right) \Psi_{t+1}^{Z} \right] \right\} \end{aligned} \tag{30}$$

where $\Psi^Z < 0$ and $\Psi^{Z'} > 0$ denote, respectively, the partial derivative of hours worked with respect to the stock of organizational capital in current period and stock of organizational capital in next period. mrs_t denotes the marginal rate of substitution between consumption and leisure $\left(mrs_t = \frac{U_{n,t}}{U_{c,t}}\right)$.

Equation (25) is standard, which states that in the case of flexible wages, workers are always on their labor supply schedule, and therefore the real wage coincide with the marginal rate of substitution between consumption and leisure. (26) is the standard Euler equation for the accumulation of physical capital which states that, at the optimum, the utility cost of sacrificing one unit of consumption must be equal to discounted utility benefit of this unit consumption tomorrow. Equation (27) indicates that a representative firm chooses the relative price for its differentiated product as a constant markup over the real marginal cost. Note that the pricing policy expressed by (27) stems from the imperfect competition

feature of the market and the flexible price allocation implies a constant real marginal cost $\lambda_t^{y_i} = \frac{1}{\eta}$.

Equation (28) describes the specific form of marginal cost which is given by the ratio of the real wage to the marginal product of effective labor engaged in goods production. In another words, given the wage rate, the intermediate good producer should optimally adjust the allocation of labor input in such a way that re-scaled marginal product of labor - marginal product of labor divided by the product of constant markup and the fraction of labor input in good production - equals the desired marginal rate of substitution between consumption and leisure. Similarly, (29) states that a intermediate goods producer should choose the utilization rate of physical capital in such a way that the ratio of the rental rate of capital to the re-scaled marginal product of capital equals the constant marginal cost. The optimality conditions, (28)-(29), distinct the costly model from the byproduct model in terms of the nature of learning mechanism. It states that if some economic prices were paid on the stock of organizational capital, the producers should explicitly take into account the potential contribution of labor and capital input on the cumulation of organizational capital in order to correctly re-scale the marginal product of factor inputs.

The term on the left hand side of (30) can be thought of as the costs of producing an additional unit of future organizational capital at period t. Let $\xi_t^w \equiv \mathcal{B}_t mrs_t - u_t^n w_t$ denotes the wage markups due to organizational capital. In the standard dynamic general equilibrium models without costly learning, real wage coincides with the marginal product of labor, and therefore $\xi_t^w = 0$. By contrast, in costly model ξ_t^w captures the extra labor cost paid for producing a unit of organizational capital, which equals the total labor cost $\mathcal{B}_t mrs_t$ less the labor cost spent on goods production $u_t^n w_t$. On the right hand side of (30), the intermediate good producer calculates the discounted benefits of this additional unit of organizational capital in next period. The first term represents $\left[\frac{D_{t+1}}{D_t}\left(\lambda_{t+1}^{y_i}\varepsilon\frac{y_{i,t+1}}{Z_{i,t+1}}\right)\right]$ units of discounted revenue attributed to an extra unit of organizational capital at period $t+1$, and the second term states that an extra unit of organizational capital at period $t+1$ also reduces Ψ_{t+1}^Z units of labor supply

and therefore lowers the discounted cost by $\left[\frac{D_{t+1}}{D_t}\left(\mathcal{B}_t mrs_t - u_t^n w_t\right)\Psi_{t+1}^Z\right]$ units at period $t+1$. In sum, the condition (30) implies that the cost of producing an additional unit of organizational capital today is equal to the discounted benefit of this organizational capital tomorrow.

- the Byprodct Model

We can also rewrite (21) as $y_t(i) = \mathrm{Y}(Z_t, Z_{t+1})$ and obtain the following optimality conditions for the byproduct model:

$$w_t = \mathcal{B}_t mrs_t \tag{31}$$

$$U_{c,t} = \beta E_t\left[U_{c,t+1}\left(r_t^k + 1 - \delta\right)\right] \tag{32}$$

$$w_t = mc_t \alpha \frac{y_{i,t}}{N_{i,t}} \tag{33}$$

$$r_t^k = mc_t(1-\alpha)\frac{y_{i,t}}{K_{i,t}} \tag{34}$$

$$v_{i,t} = \eta mc_t - \eta \frac{\lambda_t^{Z_i}}{\mathrm{Y}_t^{Z'}} \tag{35}$$

$$\left(\frac{\eta mc_t - v_{i,t}}{\eta}\right)\mathrm{Y}_t^{Z'} = E_t\left\{\frac{D_{t+1}}{D_t}\left[mc_{t+1}\varepsilon\frac{y_{i,t+1}}{Z_{i,t+1}} - \left(\frac{\eta mc_{t+1} - v_{i,t+1}}{\eta}\right)\mathrm{Y}_{t+1}^Z\right]\right\} \tag{36}$$

where $\mathrm{Y}^Z < 0$ and $\mathrm{Y}^{Z'} > 0$ denote, respectively, the partial derivative of output with respect to the stock of organizational capital in current period and stock of organizational capital in next period. Equation (31)-(34) are standard optimality conditions similar to those in the benchmark model, while (35) and (36) deserve some comments. (35) states that in setting the relative price of intermediate goods the producers solve an optimal time-varying markup problem. The term, $\eta\frac{\lambda_t^{Z_i}}{\mathrm{Y}_t^{Z'}}$ can be thought of as deviations from the standard pricing equation, which will not appear in the standard model of monopolistic competition without learning-by-doing. We rewrite (35) as following:

$$\lambda_t^{Z_i} = \left(\frac{\eta mc_t - v_{i,t}}{\eta}\right)\mathrm{Y}_t^{Z'}. \tag{35'}$$

The condition (35′) indicates that the marginal value of organizational capital is the product of two components, $\left[\frac{\eta mc_t - v_{i,t}}{\eta}\right]$ the normalized deviations from relative price by the constant markup and $Y_t^{Z'}$ units of output required to generate an additional unit of organizational capital in one period later. The first order condition (36) describes the dynamic feature of the marginal value of organizational capital. (35′) and (36) together deliver the persistent deviations from the standard relative price condition and producers thus consider the tradeoff between demand for their differentiated goods and the future productivity intratemporally and intertemporally.

3 Empirical method

In this section we discuss our methodology for estimating and evaluating the empirical performance of three competing models. We make use of Bayesian methods which have been applied to various economics literature, especially in DSGE modeling. Note that the equilibrium system of a DSGE model can be linearly approximated around its stationary steady-state in the form of

$$\mathrm{A}E_t(\hat{\mathbf{x}}_t|\mathrm{I}_t) = \mathrm{B}\hat{\mathbf{x}}_t + \mathrm{C}(F)E(\varepsilon_t|\mathrm{I}_t) \tag{37}$$

where $\hat{\mathbf{x}}_t$ is a vector of endogenous variables[4], $E_t(\hat{\mathbf{x}}_t|\mathrm{I}_t)$ is the expectation of $\hat{\mathbf{x}}_{t+1}$ given period t information, ε_t is a vector of exogenous stochastic process underlying the system, and $\mathrm{C}(F)$ is a matrix polynomial of the forward operator F. The solution of log-linearized system (37) can be written in the following state-space form:

$$\hat{s}_{t+1} = \mathrm{P}\hat{s}_t + \mathrm{C}_1\varepsilon_{t+1} \tag{38}$$

$$\hat{y}_t = \mathrm{Q}\hat{s}_t \tag{39}$$

where the vector $\varepsilon = \begin{bmatrix} \hat{\varepsilon}_{At} \\ \hat{\varepsilon}_{pt} \end{bmatrix}$ contains technology and preference innovations. Then we update the state-form solution by adding a set of measurement equation which links the observed time series to the vector

[4]For any stationary variable x_t, we define $\hat{x}_t = \left(\frac{x_t - \bar{x}}{\bar{x}}\right)$ as the percentage deviation from its steady-state value, $\bar{x}$.

of unobserved state variables. We further use the Kalman filter to evaluate the likelihood function of the state-space form solution and combine the likelihood function with our specified prior knowledge about these deep parameters to form the posterior distribution function[5]. The sequence of posterior draws can be obtained using Markov Chain Monte Carlo (MCMC) methods. We use the random-walk Metropolis-Hasting algorithm as described in Schorfheide (2000) to numerically generate the Markov chains for the structural parameters. Point estimates of Θ can be obtained from calculating the sample mean or median from the simulated Markov chains. Similarly, inference of Θ are derived from computing the percentiles of these posterior draws. Furthermore, given the sequence of posterior draws of Θ, we compute posterior statistics of interest, which are often used to validate the model performance, such as impulse response function, forecast error decomposition (FEVD) and historical decomposition.

3.1 Data

The data used in this study are drawn from the Federal Reserve Bank of St. Louis FRED website. The data sample consists of seasonally adjusted US quarterly time series, from 1954:I to 1997:IV, on total hours for non-agricultural industries and growth rate of real GDP in chained 2000 dollars. Both series are expressed in per capita terms by dividing by the civilian non-institutional population, ages 16 and over.

3.2 Prior Specification

Table 1 presents the marginal prior distributions for the structural parameters. The choice of prior distributions for parameters reflect restrictions on their natural domain, such as non-negativity or interval restrictions. Note that the priors on the structural parameters are assumed to be independent of each other, which allows for easier construction of the joint prior density used in the MCMC algorithm. Thus, the joint distribution is

[5]We describe the computational steps in appendix

assumed to the product of independent prior distributions with

$$p(\Theta|\mathcal{M}_i) = p(\alpha|\mathcal{M}_i)p(\alpha_1|\mathcal{M}_i)p(\eta|\mathcal{M}_i)\dots p(\omega_n|\mathcal{M}_i) \tag{40}$$

The depreciation rate of capital δ is assumed to follow a Beta distribution with a mean of 0.025 and standard error of 0.003. The prior for α, the labor share of nation income is described by a Beta distribution with a mean of 0.66 and standard error of 0.05. In costly LBD and C-J LBD models, we adopt the estimate for ε in Cooper and Johri (2002) as prior mean and choose 0.05 as prior standard deviation. we assume the same Beta distribution for decay rate of organizational capital, γ, across LBD models with a mean equal to 0.6 and the standard deviation of 0.05.

TABLE 1: Prior Distributions for the Structural Parameters

Parameter	**Range**	**Density**	**Mean**	**S.D.**
Learning-by-Doing Parameters, Prior 1:				
ε	$\Re$	Normal	0.1	0.05
γ	$\Re$	Normal	0.5	0.05
Learning-by-Doing Parameters, Prior 2:				
ε	$\Re$	Normal	0.1	2.5
γ	$\Re$	Normal	0.5	2.5
Additional Parameters:				
α	$[0,1]$	Beta	0.6	0.05
α_1	$[0,1]$	Beta	0.65	0.003
γ_a	$\Re$	Normal	0.005	0.005
δ	$[0,1]$	Beta	0.025	0.003
ϕ	$\Re^+$	Gamma	2	0.5
ρ_p	$[0,1]$	Beta	0.8	0.1
σ_A	$\Re^+$	Inverse Gamma	0.02	∞
σ_p	$\Re^+$	Inverse Gamma	0.02	∞

Calibrated Parameters:		
β	the discount factor	0.99
η	the price markup over marginal cost	1.1

Regarding the labor supply elasticity, we assume ϕ_n follows a Gamma distribution with a mean of 2 with a standard error of 0.5. The autocorrelation ρ_p of the preference process follows a Beta distribution with mean of 0.8 and standard deviation of 0.1. Uninformative inverse gamma distributions are used for the precision of the shocks, $\{\sigma_A, \sigma_p\}$.

As these deep parameters are largely in line with the literature, we use tight priors[6] to make the estimated model priori comparable to those in the literature. In all models, we calibrate two parameters, the discount factor and the steady-state markup of price over marginal cost. The discount factor β is set to 0.99, which implies a steady-state quarterly real interest rate of 4 per cent. For the steady-state markup, we calibrate it equal to 1.1, which implies that the elasticity of substitution between goods, $\frac{\eta}{\eta-1}$, equals 11. This value is consistent with previous literature. Chari, Kehoe and McGrattan(2000) who set it to 10 and Korenok and Swanson (2005) set it at 11.

3.3 Posterior Estimates

Based on 150,000 draws from two independent Markov chains, we compute the posterior mean and the 95 percent probability intervals for each of the parameters, with results reported in Table 2. Posterior estimates all appears reasonable. As we can see in the first column, parameter estimates for the benchmark model are similar to those in the literature. The labor share in output production is estimated to be 0.62. The estimated autoregressive coefficient of the preference shock equals 0.844 and the standard deviation of innovations of preference shock equals 0.004, which is of relatively the same order magnitude used in the literature on RBC. These two parameters play important roles in matching model predictions of hours worked with the actual aggregate hours worked series.

Of special interest here are the learning-by-doing parameters. For ε, the estimates of Byproduct and Costly models are, respectively, 0.22 and 0.19. Both estimates implies that learning rate is less than 18 percent, which are also close to the estimate by Johri and Letendre (2007) using US aggregate data. For γ, Costly model has higher posterior estimate $\gamma = 0.52$ than Byproduct does $\gamma = 0.46$. These values are also consistent with the estimates of Cooper and Johri (2002) using manufacturing data, while Johri and Letendre (2007) estimate for γ is as high as 0.8. For the

[6]The prior variance were chosen to reflect a reasonable degree of uncertainty over the calibrated values of parameters.

other parameters, posterior estimates are very similar between Costly and Byproduct models. It is worth noting that LBD models require slightly less persistent preference shock to fit the time series than the benchmark model does since the learning-by-doing provides an internal propagation mechanism.

TABLE 2: Posterior Estimates for the Structural Parameters

	Baseline		**By-product LBD**		**Costly LBD**	
Parameter	Post Mean	NSE	Post Mean	NSE	Post Mean	NSE
α	0.616	0.015	0.655	0.021	0.632	0.021
ε	-	-	0.221	0.012	0.194	0.012
γ_a	0.003	0.007	0.003	0.007	0.003	0.007
γ	-	-	0.462	0.039	0.521	0.038
δ	0.022	0.003	0.020	0.003	0.020	0.003
ϕ	0.752	0.040	1.062	0.040	0.954	0.040
α_1	-	-	-	-	0.653	0.037
ρ_p	0.844	0.011	0.842	0.011	0.838	0.011
σ_A	0.012	0.008	0.011	0.009	0.013	0.008
σ_p	0.004	0.003	0.003	0.003	0.003	0.003

Notes: The posterior means are calculated from the output of the Metropolis-Hastings algorithm. NSE is the numerical standard error.

The Bayesian approach also allows for the explicit evaluation of model uncertainty. We conduct formal comparison of overall time series fit between three non-nested hypothetical DSGE models and report the marginal data densities and posterior odds ratios in Table 3. The posterior odds ratios of LBD specification versus benchmark stochastic-growth model clearly indicate that LBD improve the time series fit of DSGE model. The results suggest that in order to choose Benchmark model over Byproduct and Costly model, we need a prior probability over Benchmark model 1.10×10^{18} and 6.61×10^{16} times larger than our prior probability over Byproduct and Costly model, respectively. As for LBD models, the posterior odds ratio of Byproduct versus Costly specification indicates Byproduct model is more favorable by the data. However, we need only a prior probability over Costly model 16 times larger than our prior over Byproduct model in order to choose Costly model. As this factor is not large enough, Byproduct model outperforms the Costly model only by a slim margin. Finally note that the time-series fit of all models are worse

than that of VAR(4).

TABLE 3: Goodness of Fit

Statistic	Benchmark	Byproduct	Costly	VAR(4)
Prior probability, $\pi_{i,0}$	1/4	1/4	1/4	1/4
Log marginal data density	1036.05	1077.59	1074.78	1082.72
Posterior probability, $\pi_{i,T}$	0	0.006	0	0.994
Posterior odds ratio	1.00	1.10×10^{18}	6.61×10^{16}	1.86×10^{20}

3.4 Impulse-Response Dynamics

To shed more light on how well the DSGE models capture the dynamics of output growth and hours worked and how LBD specifications closely equivalent to each other, we examine the implied impulse-response functions and second-order unconditional moments of interest. Note that in our analysis, the model economy are driven by a random-walk technology and a stationary preference shock. The innovations in the technology process have a permanent effect on output whereas the innovations in preference process have a transitory effect. To make fair comparison between DSGE models and the a-theoretical VAR model, we employ Blanchard and Quah's (1989) method to identify the permanent and transitory shocks in the VAR.

The first column of Figure 1 depicts the posterior means of the impulse-response of output and hours worked to a one-standard deviation of permanent shock, generated by the Benchmark, Costly, by-product and VAR models. As we can see, in response to technology shocks, benchmark and learning models generate completely different patterns. In LBD models, both output growth and hours worked display inertial response to the technology shocks. Over the short horizons, impulse-responses from learning models track the VAR-based counterpart more closely than benchmark model. The second column of Figure 1 reports the posterior means of the impulse response to a one-standard deviation of transitory shocks for each model. As documented by the previous literature, the

VAR responses of output to the transitory shock exhibit a pronounced hump-shape and trend reverting path. Benchmark model fails to generate an important trend-reverting component in output, while both Costly and Byproduct models produce a pronounced hump-shaped output response, which matches the VAR response fairly well. In the response of hours worked to a transitory shock, both learning models generate observational equivalent paths, which display monotonic convergence of hours towards its steady-state. Although the responses of hours predicted by the two learning models do not fit into the confidence interval of VAR-based response over the longer horizons, they do match the shape and the magnitude of their VAR counterpart.

It is natural to ask what is the underlying driving force for learning models such that they clearly discriminate themselves from the benchmark model and why learning-by-doing can make the output and hours worked display hump-shaped response to exogenous shocks, exactly as predicted by the VAR. We explore these questions in the following sections.

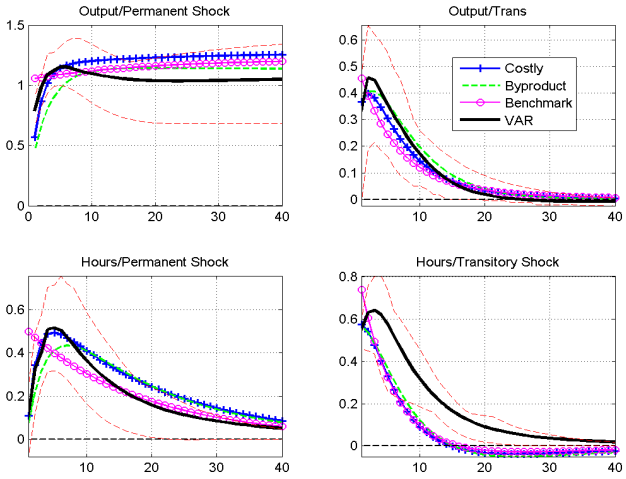

FIG. 1: Impulse Response Function (posterior mean)

3.5 Roles of Learning-by-doing in Propagating Shocks

Consider the effect of a one-standard deviation shock to productivity in (12), which is essentially a one-time shock to the growth rate of productivity, γ_a. The specification of the productivity process in (12) implies that any shock to productivity will have permanent effects. Given the random walk process in (12) we assume zero persistence of a deviation of the growth rate from its steady state level. In the period after the shock, the level of productivity is pushed up by the shock but no further changes in productivity will take place. Hence, a permanent increase in the level of productivity implies that both input factors in benchmark model, labor and physical capital can be used more efficiently. Higher productivity means higher factor prices and thus incomes from labor and capital increase as well. A higher rental rate of capital stimulates investment in the impact period, and consequently the capital stock will rise. Since physical capital is the only endogenous state variable in the benchmark model, the permanent increase in capital stock will lead to permanent increase in other variables. Notice that the graph on the left shoulder of Figure 1 plots the percentage deviation of logarithms of output from its pre-shock steady state level for the respective model. The circled line is the impulse responses from the benchmark model to a one-standard deviation of permanent shock. Output rises over 0.9 percent on the impact and then converges smoothly to its new steady-state. The graph on left-bottom of Figure 1 displays the percentage deviation of log level of hours worked from its steady-state level. If again focusing on the impulse responses from the benchmark model first, we find an interesting result that the increase in income does not induce agents to cut back their labor supply. It is the higher wage rate that makes individuals rather to supply more labor in equilibrium to take advantage of increase in productivity in the impact period.

In contrast to the benchmark model, costly and byproduct models introduce learning-by-doing which leads to persistence in the adjustment of productivity. As shown in Figure 2, because of the impact of learning-by-doing, a shock to productivity is not just a one-time shock of the growth

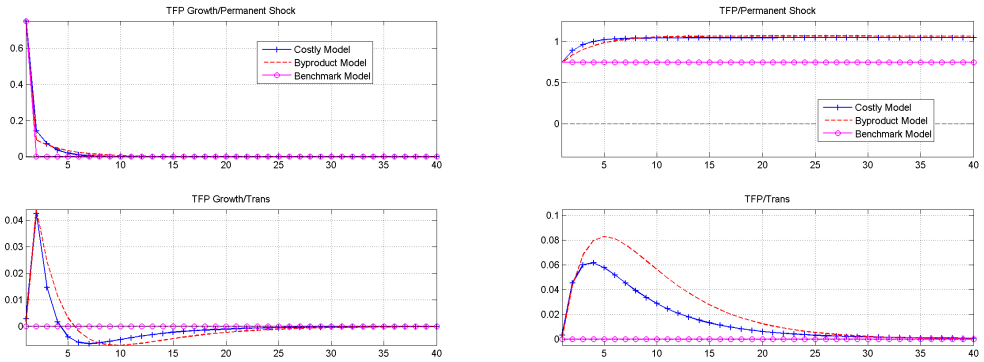

FIG. 2: Impulse Response (posterior mean): TFP

rate of productivity. It rather implies a temporary increase in growth rate of endogenous productivity which pushes up even higher the future level of productivity. In the first column of Figure 1, the impulse response dynamics for costly and byproduct models clearly illustrates the consequence in case of introducing learning-by-doing mechanism. The "hump-shaped" responses of output and hours worked display totally different patterns than the benchmark model does. Agents anticipate that productivity will be higher in the future other than on the impact period hence they initially cut back hours worked and enjoy more leisure. This is optimal because along the new balance equilibrium path, labor will be more productive than it is now. Therefore it makes more sense to consume leisure now and work harder in the future.

The second column of Figure 1 reveals a similar pattern in the case of a shock to preference. In the impact period, the response of output and hours worked are quiet similar in these three models. Thereafter, the impulse dynamics are different. In the benchmark model, output and hours worked smoothly declines to their original steady-state. Figure 2 shows that the total factor productivity in benchmark model is unaffected by this preference shock at all. In costly and byproduct models, however, the k-period-ahead effect of a preference shock is larger than the impact effect and hence output display hump-shaped responses to the preference shock but not for the hours worked in these model. Figure 2 clearly illustrates that learning-by-doing transforms a shock to preference into a tem-

porary increase in growth rate of productivity. The accumulative effects of temporary productivity increase imply that future productivity will be even higher, which induces hump-shaped responses of hours worked and investment. As a consequence, the hump-shaped response of output to preference shocks can be attributed to its own factor inputs dynamics.

3.6 Costly vs. Byproduct Hypothesis

In what follows, we briefly comment on the robustness of our preliminary estimation results to the way that we model learning-by-doing mechanism. We find that the costly and byproduct models have quite similar qualitative and quantitative implications for aggregate variables to different types of shocks. Given the posterior estimates, the response adjustment of output, hours worked and total factor productivity display quite similar patterns across the two models. Figure (3) also plots the response

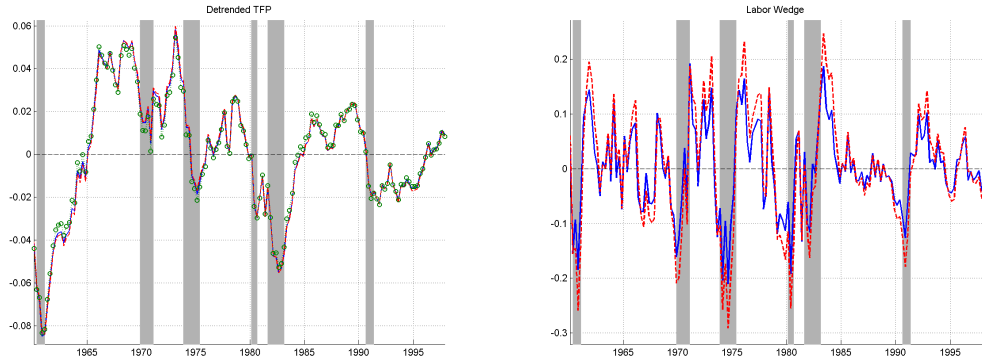

FIG. 3: Smoothed TFP and Labor Wedge: Circled line refers to Benchmark model, solid blue lines to Costly model, dash lines to Byproduct model.

of markup in byproduct model against the response of labor share and capital share in costly model. To conserve on space, we do not report the impulse-response functions for all the other variables in the two learning models; but they are all quantitatively and qualitatively similar to each other. This finding is robust irrespective of a shock to technology or to preference. The main difference, however, is that costly model requires slightly more volatile technology innovation than the byproduct model

does. The standard deviations of technology innovations are 1.12% and 1.25% for costly and byproduct model, respectively, while both learning models give us the same magnitude of the standard deviation of preference shocks. Interestingly, it is because of this difference between estimated technology shocks across two learning models that makes two models generate observational equivalent smoothed total factor productivity. As shown on the left panel of Figure 3, the detrended total factor productivity over the post war period, costly and byproduct model have quite similar patterns. The shaded vertical areas correspond to the official recession periods according to NBER. It is worth noting that the total factor productivity from two learning models consist of the technology shock A_t and the organizational capital Z_t, while A_t is the only component of the total factor productivity from the benchmark model. Another

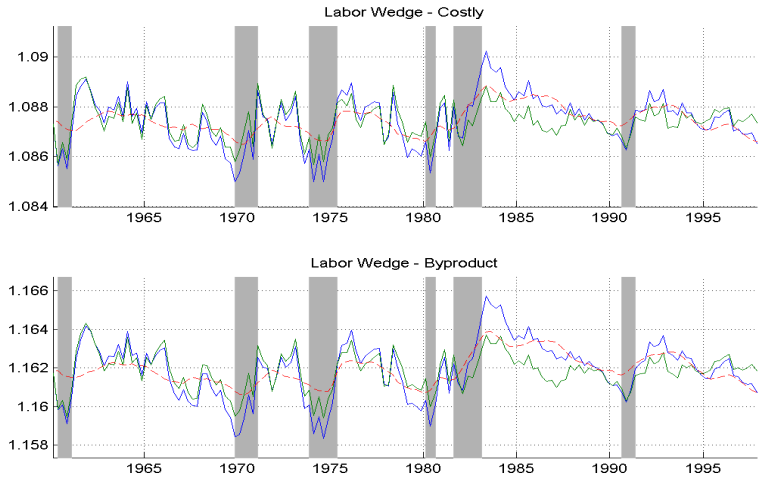

FIG. 4: Labor Wedge Decompositions: Solid blue lines refer to the smoothed series, solid green lines to the recovered series with technology shocks, dash lines to the recovered series with preference shocks.

way to compare the costly model with the byproduct model is using the respective labor market conditions implied by these two learning models. We define the labor wedge $Wedge_t = \frac{w_t}{MPL_t}$, as the ratio of the real wage and marginal product of labor (MPL). The labor wedges in the byproduct

and costly model are, respectively,

$$\begin{aligned} Wedge_t^{bypd} &= mc_t \\ Wedge_t^{cost} &= \frac{1}{\nu u_t^n} \end{aligned}$$

Both learning models drive a wedge between labor productivity and wages but through different channels. The by-product model delivers an endogenous time varying price-cost markup, while the costly model drives a time varying wedge between the wage and the marginal product of labor. The right panel of Figure3 describes the behavior of the smoothed labor wedges over the whole thirty-eight year period. In addition, unlike the benchmark model, where the labor wedge is constant, the learning-by-doing models have their labor wedges to be a combination of technology shocks and preference shocks. In Figure 4, we provide an exact decompositions of the labor wedge implied from two learning models into the components driven by the smoothed technology and preference shocks. As can be seen here, costly model and byproduct model also display quite similar pattern along these dimensions. In the next section, we will compute formal statistics to estimated and evaluate different models of interest.

3.7 Shock Decompositions

3.7.1 Variance decomposition

Table 2 displays the results of these forecast error variance decompositions, which give us the fractions of the observed output growth and hours worked in the US economy are explained by the DSGE models. Given the technology is modeled as a random walk process, the technology innovations account for a very large share of the unconditional variance in aggregate output due to the cumulation effects of technology shocks. The first column of the top panel suggests that technology shock accounts for about 84% of the one-period-ahead forecast error variance of output growth and its contribution to output growth monotonically decreases along the time horizon. The last line of panel A, $k = \infty$ indicates that the technology shock can account for 83% of the unconditional

variance in output growth. In contrast to benchmark model, costly and byproduct models display different patterns in terms of the contribution of technology shocks to output growth, but both models have quite similar pattern. The technology shock accounts for 83.2% and 81.6% of the one-quarter-ahead forecast error variance in output growth, respectively, for byproduct and costly model. On the other hand, as shown in Table 2, the technology shock accounts even more for the k-step-ahead forecast error variances for values of k ranging from 4 to 40 quarters. These results suggest that learning-by-doing plays an important role in explaining output fluctuations over the business cycle frequencies.

In addition, panel B of Table 2 contains interesting results. DSGE models with learning-by-doing give us quite different predictions for hours worked from the benchmark model. In benchmark model, the technology shock explains 60% of the unconditional variance in hours, and it only explains more than 31% of the one-quarter-ahead forecast error in hours. Note that Christiano and Eichenbaum (1990) show that the model's predictions change in an important way when the technology process changes from a random walk to a stationary AR(1). Because of the accumulative effects of random work technology, our result for benchmark model is largely different from the previous findings in the literature where AR(1) technology process is employed. Using stationary technology process, Ireland (2004) find that technology shocks accounts for almost none of the unconditional variance of hours worked.

Both byproduct and costly models generate surprising results for hours worked. As shown in panel B, technology shocks accounts for merely 3.6% and 1.9%, respectively, of the the one-quarter-ahead forecast error variance in the hours series, although the technology shocks in byproduct and costly model, respectively, explain more than 72% and 65% of the unconditional variance in hours worked. These results are consistent with the posterior impulse-response of hours worked which displays inertial response to the technology shock for both LBD models. These results are also in line with the previous findings reported by Watson (1993). Watson (1993), in particular, documents that for the key aggregate variables including hours worked, the spectral power is significant less in

very low frequency than in business cycle frequency.

TABLE 4: FEVD: Percentage of variance due to technology

Quarter Ahead	Benchmark	Byproduct	Costly
(A) Output Growth			
1	84.180	83.237	81.614
4	83.558	85.217	82.974
8	83.297	84.748	82.759
16	83.227	84.559	82.455
20	83.201	84.493	82.294
40	83.198	84.488	82.279
∞	83.198	84.488	82.279
(B) Hours worked			
1	31.185	3.550	1.926
4	38.970	36.162	22.390
8	46.894	54.177	42.736
16	51.919	62.373	53.745
20	56.606	68.770	62.363
40	58.984	71.677	65.609
∞	59.259	72.024	65.876

3.7.2 Historical Decomposition

Figure 5 illustrates that DSGE models nearly explain 100% of the the variation in output growth and hours worked and thus provide a more insightful historical decomposition. Specifically, we can compute the underlying structural shocks using Kalman filter, taking the estimated parameters as given. To measure the contribution of technology or preference shock to a given variable, we shut down one of the shocks and simulated the model. We then obtain paths of output growth and hours, which would have taken place if only technology or preference shocks are present. It helps us to compare the actual data series to their hypothetical series where only one of the shocks occurs. Figure summarize the historical contribution of the structural shocks to output growth and hours worked. The shaded vertical areas correspond to the official recession periods according to NBER.

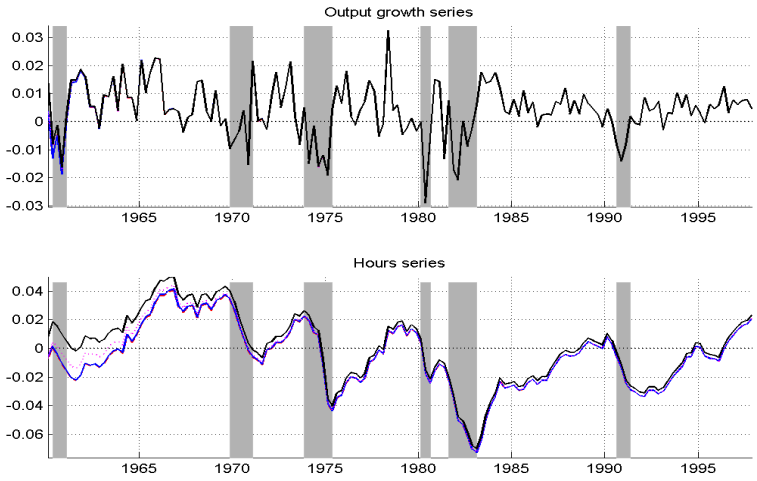

FIG. 5: Smoothed Observable Variables: Solid black lines refer to the data, dotted lines to the benchmark model, solid blue lines to Costly model, dash lines to byproduct model.

Focusing on the historical decomposition of output growth first, it is clear that across three DSGE models, the short-run variability in output growth is mostly accounted by technology shocks, which is in line with the results from the variance decomposition. Our decomposition results suggest that output growth is mainly driven by technology shocks in recession period. Such patterns are similar across three models.

For hours worked, the historical decompositions are also consistent with the forecast error variance decomposition results. By contrast, preference shock now play an more important role in explaining historical variability of hours work. According to our model, preference shock was the predominant factor behind the drops in hours occurring in mid 1970s, throughout 1980s and afterwards, while technology shocks contributed only moderately to hours variation in recession periods and it was the key factor behind the surge in hours worked in 1960s.

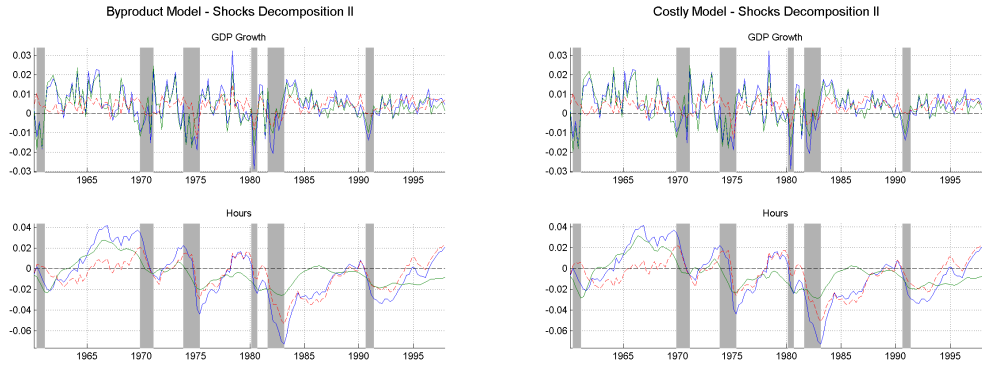

FIG. 6: Historical Decompositions: Solid blue lines refer to the smoothed series, solid green lines to the recovered series with technology shocks, dash lines to the recovered series with preference shocks.

3.8 Moments of Interest

3.8.1 Persistence of Output Growth

An important shortcoming of standard RBC model is that it lacks the endogenous propagation mechanism to generate enough persistence in the endogenous variables when facing exogenous shocks (e.g. Cogley and Nason, 1995), while many previous studies find that output growth is positively autocorrelated over short horizons and weakly autocorrelated over longer horizons (e.g., Cochrane, 1998 and Chang, Gomes and Schorfheide, 2002).

In Table 5, we compare the predicted autocorrelations of output and hours worked of the benchmark and LBD models to US data for the period 1960:1 to 1997:4. In panel A of Table 5, the results clearly show that the benchmark model predicts the autocorrelations of output growth to be essentially zero, while the costly and byproduct model are capable of generating positive autocorrelations of output growth, which match autocorrelations in the data quite well over the short horizons. In order to formally evaluate models using model-based and observed autocorrelations, we specify two posterior expected loss functions, L_q and L_{χ^2}.[7] Both measures of loss reported in panel A confirm that: 1) learning-by-doing

[7]See Schorfheide (2000) for a detailed discussion of these loss functions and their interpretations

models does much better than the benchmark model in explaining output growth autocorrelations; 2) costly and byproduct models are barely discriminated by the results based on these loss functions.

Panel B in Table 5 reports the autocorrelations of hours worked predicted by the three models. A lag-by-lag comparison indicates that both costly model and byproduct models are successful in replicating he sample autocorrelation and are closely equivalent to each other. The loss statistics suggest that byproduct model does marginally better than the costly model.

TABLE 5: Autocorrelation Statistics

Statistic	Lag	Benchmark	Byproduct	Costly	VAR
(A) *Output Growth*, Corr(Δ ln GDP,Δ ln GDP(-j)):					
Posterior Mean	1	0.002	0.194	0.325	0.322 [0.217,0.462]
	2	0.085	0.145	0.157	0.207 [0.189,0.462]
	3	0.019	0.031	-0.004	0.067 [-0.026,0.170]
	4	-0.025	0.059	-0.029	-0.026 [-0.0362, -0.009]
L_q risk	1-4	0.121	0.040	0.017	
L_{χ^2} risk	1-4	0.982	0.006	0.016	
(B) *Hours worked*, Corr(ln N, ln N(-j)):					
Posterior Mean	1	0.926	0.947	0.948	0.971 [0.962,0.980]
	2	0.852	0.875	0.864	0.908 [0.883,0.939]
	3	0.780	0.798	0.784	0.825 [0.787,0.886]
	4	0.706	0.731	0.712	0.734 [0.692,0.828]
L_q risk	1-4	0.008	0.003	0.005	
L_{χ^2} risk	1-4	0.183	0.113	0.119	

3.8.2 Other Second-order Unconditional Moments

Traditional way to validate DSGE models is to check out their ability to match a fairly comprehensive set of stylized facts from data. Table 6 reports the model-based second-order unconditional moments as well as those in the data. These business cycle statistics from the data are obtained by using Hodrick-Prescott filtered aggregate time series and the statistics from the estimated model are obtained by using Hodrick-Prescott filtered smoothed time series from the models.

The estimated byproduct and costly models provide a good match on most dimensions of the data and again both models give us closely equivalent second-order moments along every single dimension. It is worth noting that these learning-by-doing models have interesting implications for the dynamics of the labor market. During the postwar period, two business cycle stylized facts are documented in the literature: 1) hours worked are more volatile than the average labor productivity (Kydland and Prescott, 1982); and 2) No significant correlation exits between hours and real wages. The learning-by-doing models both account very well for these facts in labor market. First, costly and byproduct models predict, respectively, the volatility of hours worked are 1.51 and 1.52 times larger than that of labor productivity, compared with 1.76 times larger in the data. Second, correlation between hours worked and real wage in costly and byproduct models are, respectively, 0.105 and 0.058, while it is -0.053 in the data and 0.263 according to the benchmark model.

Standard RBC Models where technology shocks play a major role usually generate highly procyclical real wages. By contrast, learning-by-doing models can predict mildly procyclical real wages, though technology shocks are the main driving force in the economy. Costly and byproduct models predict that the correlation between real wage and output are, respectively, 0.518 and 0.561, which is close to the correlation of 0.372 according to the data.

There are still some dimensions that predictions of our models fail to match those of the data. In particular, the correlation between average labor productivity and real wage in costly and byproduct model are,

respectively, 0.981 and 0.968, while it is 0.673 in the data. As expected, benchmark model predicts perfect correlation between average labor productivity and real wage.

TABLE 6: Second-Order Moments in the Benchmark and LBD Models

Moments	US Data	Benchmark	Byproduct	Costly
σ_c/σ_y	0.506	0.490	0.343	0.371
σ_i/σ_y	2.868	2.624	2.325	2.381
σ_n/σ_y	0.854	0.748	0.745	0.745
σ_w/σ_y	0.637	0.495	0.304	0.352
σ_{apl}/σ_y	0.515	0.495	0.491	0.492
σ_n/σ_{apl}	1.762	1.510	1.518	1.513
$Corr(c,y)$	0.911	0.965	0.911	0.915
$Corr(i,y)$	0.963	0.989	0.929	0.952
$Corr(n,y)$	0.819	0.878	0.882	0.881
$Corr(w,y)$	0.372	0.692	0.518	0.561
$Corr(apl,y)$	0.519	0.692	0.698	0.699
$Corr(apl,w)$	0.673	1.000	0.968	0.982
$Corr(n,w)$	-0.012	0.263	0.058	0.105
$Corr(n,apl)$	-0.054	0.263	0.278	0.277

3.9 Sensitivity Analysis

3.9.1 A Check with Defuse Prior Distributions

The results discussed in the previous section are conditional on using tight prior for the learning parameters ε and γ.

Table 7 reports posterior estimates under the defuse priors. The posterior means of ε are similar across costly and byproduct model, whereas the posterior means of γ are quite different across two learning-by-doing model. In terms of posterior estimates of other structural parameters, we have the new estimates very close to what we obtain under tight priors.

The posterior odds ratio reported in Table 8 indicates that the byproduct model is in favored over costly model by a factor of 167 to 1 in this case. Thus, the overall time-series fit of the byproduct model is still better than that of costly model under defuse prior for the learning parameters.

TABLE 7: Posterior Estimates: Defuse Prior

	Benchmark		**By-product**		**Costly**	
Parameter	Post Mean	NSE	Post Mean	NSE	Post Mean	NSE
α	0.616	0.015	0.652	0.021	0.636	0.021
ε	-	-	0.235	0.012	0.224	0.012
γ_a	0.003	0.007	0.003	0.007	0.003	0.007
γ	-	-	0.296	0.039	0.607	0.038
δ	0.022	0.003	0.021	0.003	0.020	0.003
ϕ	0.752	0.040	1.090	0.040	0.988	0.040
α_1	-	-	-	-	0.650	0.037
ρ_p	0.844	0.011	0.840	0.011	0.837	0.011
σ_A	0.012	0.008	0.011	0.009	0.013	0.008
σ_p	0.004	0.003	0.003	0.003	0.003	0.003

Figure 7 illustrates the posterior impulse-response functions of the costly and byproduct models with defuse priors. The left column of Figure 7 shows that impulse responses to a permanent technology shock. Costly and byproduct model generate observational equivalent paths for hours worked, a pronounced hump that reaches its peak 6 quarters after the shock, exactly as predicted by the VAR. In addition, the output responses generated by two learning models fit well into the confidence interval of the VAR-base counterpart. In response to a transitory shock, the right column of Figure 7 displays that impulse responses from two learning model under defuse priors. In the response of output, the costly model generates a small hump as it does under tight priors, while the byproduct model, on the other hand, produce a more pronounced hump that reaches it peak three quarters after the transitory shock and tracks the VAR response much more closely than that implied by the costly model. In the response of hours worked to a transitory shock, both learning models generate observational equivalent paths, which display monotonic convergence of hours towards its steady-state. Although the responses of hours predicted by the two learning models do not fit into the confidence interval of VAR-based response function over the longer horizons, they do match the shape and the magnitude of the VAR counterpart. Finally,

TABLE 8: Goodness of Fit: Defuse Prior

Statistic	Benchmark	Byproduct	Costly	VAR(4)
Prior probability, $\pi_{i,0}$	1/4	1/4	1/4	1/4
Log marginal data density	1036.05	1075.53	1070.41	1082.72
Posterior probability, $\pi_{i,T}$	0	0.001	0	0.999
Posterior odds ratio	1.00	1.40×10^{17}	8.36×10^{14}	1.86×10^{20}

the learning models under defuse priors generate quite similar pattern as the models under tight priors do.

In summary, our results that the byproduct model fits the data marginally better than the costly model appears to be robust to the use of defuse priors for the learning parameters ε and γ and both learning models still outperform the benchmark model by a factor of 8.36×10^{14} to 1 and over in this case.

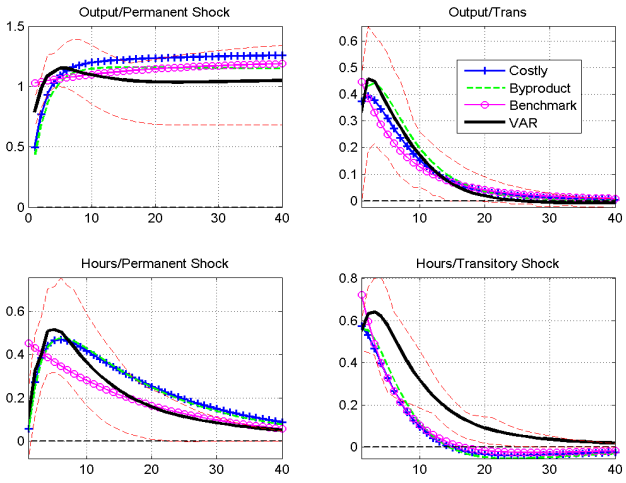

FIG. 7: Posterior IRF: Defuse Prior

3.9.2 Sensitivity Check with Learning Parameters

Above estimation results indicate that byproduct learning and costly learning models are only marginally different from each other. In order

to safely address the point that our business cycle results are robust to the way that we model learning-by-doing, costly or byproduct, we conduct further sensitivity check on two most important parameters which determine the learning dynamics behind our main results. If it turns out that both learning-by-doing models exhibit the same cyclical behavior for all possible scenarios of learning dynamics the economy would have had, then it seems that the assumptions on the types of learning mechanism, costly or byproduct, are truly innocuous. We first looked at the role of learning rate in production technology by assuming that the learning parameter ε takes several reasonable values, while the rest of of parameters in both learning-by-doing models remains the same as before. The left panel of Figure 8 shows the hypothetical impulse responses from the byproduct model to technology and preference shocks, respectively, and the right panel displays the impulse responses from the costly model. Byproduct model produces imaginary eigenvalues with $\varepsilon = 0.5$, while the costly model can take larger learning rates as high as $\varepsilon = 0.625$.

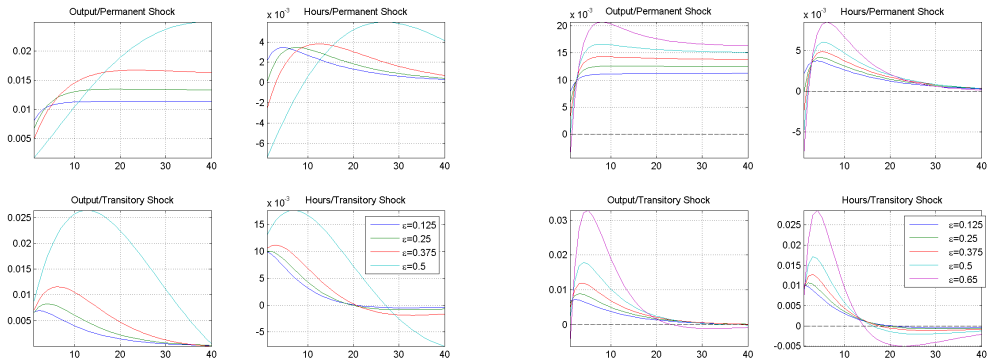

FIG. 8: Impulse Responses: Left panel refers to Byproduct model, right panel refers to Costly model

It is worth noting that as learning rate increases, both learning models predict that hours worked fall immediately following a technology shock. These result might led some light on the recent and active debate known as the "hours debate" as it centers on whether short-run response of hours worked to a positive technology shock. Gali (1999) finds that for the majority of the G7 countries, hours worked fall following a technol-

ogy shock. He estimated a VAR of the first differences of hours and labor productivity and then used Blanchard and Quah's identification strategy to identify the technology shock. Gali's discovery fail the RBC models to be a valid analytical tool for studying business cycle fluctuations. However, our results show that the DSGE model with learning can predict negative responses of hours to a technology shock in short horizons, because a high learning rate implies that total factor productivity is affected greatly by organizational capital and pushes up future level of productive even higher. Since hour productivity will be much higher than it is now, agents will cut back their hours worked even more than they would do when learning rates are low. Therefore, we obtain the negative responses of hours worked to a technology shock over short horizons. The left panel of Figure 9 shows that the responses of TFP growth from byproduct learning model to a technology shock and a preference shock, respective, and the right panel displays the responses of TFP growth from costly model to exogenous shocks. We can clearly find the pattern that the TFP growth becomes more persistent as ε increases and the technology shock has more pronounced effect on the TFP growth in costly model.

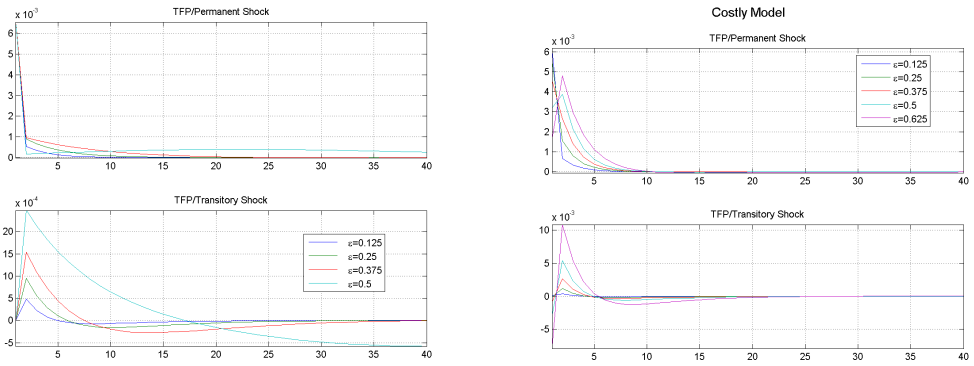

FIG. 9: Impulse Responses: Left panel refers to Byproduct model, right panel refers to Costly model

The other possible concern is whether the learning dynamics is sensitive to the values of γ so that costly learning can easily discriminate itself from byproduct learning. To save space, we do not report the impulse-response functions and second-order moments for the costly learning and

byproduct learning models. But along this dimension, two learning-by-doing models produce quite similar results.

4 Conclusion

In this paper we write down a simple DSGE model of costly learning-by-doing to address the view that is opposed to the traditional formulation of byproduct learning. We provide the aggregate estimates of both types of models and compute formal statistics to assess the robustness of our business cycle results to the way that we models learning-by-doing. We find that costly and byproduct models generate quite similar business cycle statistics. Using Bayesian techniques, however, we have found that byproduct model fits aggregate U.S. data marginally better than costly model.

Learning-by-doing leads to persistence in the adjustment of total factor productivity in response to a technology shock. Given this fact, we find that it is very likely that hours worked fall following a positive technology shock. This finding might shed some light on the recent "hours debate".

References

[1] Arrow, K.J. (1962). The Economic Implications of Learning by Doing. *Review of Economic Studies* 80, pp. 155C173.

[2] Atkeson, A. and Kehoe, P.L. (2002) Measuring Organizational Capital. *NBER working paper,* 8722.

[3] Bahk, M-H and Gort, M. (1993) Decomposing Learning by Doing in New Plants. *Journal of Political Economy* 101, (4), pp. 561-583

[4] Benkard, C. L. (2000) Learning and Forgotting: the Dynamics of Aircraft Production. *American Economic Review* 90, (4), pp. 1034-1054

[5] Blanchard, O. and Quah, D. (1989) The Dynamic Effects of Aggregate Demand and Aggregate Supply Disturbances, *American Economic Review*, 79, pp.655-673.

[6] Chang, Y., Gomes, J.F., and Schorfheide (2002) Learning-by-Doing as a Propagation Mechanism. *American Economic Review* 92, (5), pp. 1498-1920

[7] Chari, V., Kehoe, P.J. and McGrattan, E.R. (2000) Sticky Price Models of the Business Cycle: Can the Contract Multiplier Solve the Persistence Problem? *Econometrica* 68, (5), pp. 1151-1180

[8] Christiano, L. and Eichenbaum, Martin (1990) Unit Root in Real GNP: Do We Know, and Do We Care? *Carnegie-Rochester Conference Series on Public Policy* 32, pp. 7-16

[9] Cochrane, J.H. (1988) How Big Is the Random Walk in GNP? *Journal of Political Economy* 96(5), pp. 893-920

[10] Cogley, T. and Nason, J.M. (1995) Output Dynamics in Real-Business-Cycle Models. *American Economic Review* 85(3), pp. 492-511

[11] Cooper, R.W. and Johri, A. (2002) Learning-by-Doing and Aggregate Fluctuations. *Journal of Monetory Economics* 49, pp. 1539-1566

[12] Clarke, A.J. and Johri, A. (2008) Pro-cyclical Solow Residuals without Technology Shocks. *McMaster Univeristy Working paper*

[13] M. Dotsey, R.G. King and A.L. Wolman (1999) State Dependent Pricing and the General Equilibrium Dynamics of Money and Output. *Quarterly Journal of Economics* 114, pp. 665-690

[14] Gali, J. (1999) Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations? *American Economic Review* 89(1), pp. 249-271

[15] Ireland, P.N. (2004) A method for taking models to the data. *Journal of Economic Dynamics and Control* 28(6), pp. 1205-1226

[16] Irwin, D.A. and Klenow, P.J. (1994) Learning-by-Doing Spillovers in the Semiconductor Industry. *Journal of Political Economics* 102(6), pp. 1200-1227

[17] Jarmin, R.S. (1994) Learning-by-Doing and Competition in Early Rayon Industry. *RAND Journal of Economics* 25(3), pp. 441-454

[18] Johri, A. (2005) Learning-by-Doing and Endogenous Price-Level Inertia. *working paper*

[19] Johri, A. and Letendre, M-A. (2007) What do residuals from first order conditions reveal about DGE models, (with), *Journal of Economic Dynamics and Control*, 31(8), pp. 2744-2773.

[20] Kydland, F.E. and Prescott, E.C. (1982). Time to Build and Aggregate Fluctuations. *Econometrica* 50, pp. 1345-1370.

[21] Korenok, O. and Swanson, N.R. 2005. The Incremental Predictive Information Associated with Using New Keynesian DSGE Models

vs. Simple Linear Econometric Models. *Oxford Bulletin of Economics and Statistics* 67, pp. 905-930.

[22] Lev, B. and Radhakrishnan, S. (2003) The Measurement of Firm-Specific Organization Capital. *NBER working paper* 9581.

[23] Rosen, S. (1972) Learning by Experience as Joint Production. *Quarterly Journal of Economics* 86(3), pp. 366-382

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

[25] Thornton, R.A. and Thompson, P. (2000) Learning from Experience and Learning from Others: An Exploration of Learning and Spillovers in Wartime Shipbuilding. *American Economic Review* 91, (5), pp. 1350-1369

[26] Watson, M.W. (1993) Measures of Fit for Calibrated Models. *Journal of Political Economy* 101(6), pp. 1011-1040

A Appendix

A.1 Posterior Distribution and Moment

We wish to estimate a DSGE model $\mathcal{M}_i$ and its associated vector of structural parameters Θ_i. Let

$$\Theta_i = \{\alpha, \delta, \varepsilon, \gamma, \psi, \eta, \rho_p, \sigma_A, \sigma_p, \sigma_n\}'$$

We update the state-form solution (38)-(39) by adding a set of measurement equation which links the observed time series to the vector of unobserved state variables:

$$S_{t+1} = AS_t + B\varepsilon_{t+1} \tag{A-1}$$

$$Y_t = CS_t \tag{A-2}$$

where the matrices A, B and C are functions of the models' structural parameters, and C represents the relationship between the observed data Y_t and variables in state equation S_t. $S_t = \{\hat{x}_t\}$ from equations (38) and Y_t contains only two observed control variables in $\{\hat{y}_t\}$ from equation (39). Specifically, Y_t is a 2×1 vector of observable variables, including GDP growth and hours worked; ε_t is the vector containing technology and preference innovations.[8]. Given the state-space form defined by (A-1) - (A-2), the likelihood function of the model $\mathcal{M}_i$, can be constructed by applying the Kalman filter as outlined by Hamilton (1994):

$$\ln \mathcal{L}(\Theta|Y^T, \mathcal{M}_i) = -\frac{nT}{2}\ln 2\pi - \sum_{t-1}^{T}\left[\frac{1}{2}\ln|\Omega_{t|t-1}| + \frac{1}{2}\omega_t'\Omega_{t|t-1}^{-1}\omega_t\right] \tag{A-3}$$

where the vector Θ_i contains the parameters to be estimated; $\{\omega_t\}_{t=1}^{T}$[9] is a series of innovations that are used to evaluate the likelihood function $\mathcal{M}_i$ for the data sample, Y^T, and $\Omega_{t|t-1} = E\omega_t\omega_t'$ is the variance-covariance matrix that depends on the structural parameters, Θ_i.

[8]Note that in contrast to Ireland (2004), we do not specify the measurement errors in measurement equations.

[9]ω_t is defined as $\omega_t = y_t - \hat{y}_{t|t-1}$ and $\omega_t \sim N(0, \Omega_{t|t-1})$ is assumed normally distributed

We further combine the likelihood function with our specified prior knowledge about these deep parameters to form the posterior distribution function. In the Bayesian context, the posterior distribution of Θ_i can be thought of as a way of weighting the likelihood information contained in the observed data by the prior density $p(\Theta_i|\mathcal{M}_i)$. Given a prior, the posterior density kernel[10] of Θ_i can be written as:

$$p(\Theta|Y^T, \mathcal{M}_i) \propto \mathcal{L}(Y^T|\Theta, \mathcal{M}_i)p(\Theta|\mathcal{M}_i) \tag{A-4}$$

where $\mathcal{L}(Y^T|\Theta, \mathcal{M}_i)$ is the likelihood conditional on the observed data, $Y^T = \{y_1, \ldots, y_T\}_{t=1}^T$. The sequence of posterior draws can be obtained using Markov Chain Monte Carlo (MCMC) methods. We use the random-walk Metropolis-Hasting algorithm as described in Schorfheide (2000) to numerically generate the Markov chains for the structural parameters. Point estimates of Θ_i can be obtained from calculating the sample mean or median from the simulated Markov chains. Similarly, inference of Θ_i are derived from computing the percentiles of these posterior draws.

Furthermore, given the sequence of posterior draws, $\{\Theta_i^j\}_{j=1}^N \sim p(\Theta_i|Y^T, \mathcal{M}_i)$, by the law of large numbers:

$$E\left(g(\Theta_i)|Y^T\right) = \frac{1}{N}\sum_{j=1}^{N} g(\Theta_i^j) \tag{A-5}$$

where $g(\cdot)$ is some function of interest, such as impulse response functions and moments. We can employ Markov chain Monte Carlo (MCMC) methods to evaluate equation (A-5) with $\{\Theta_i^j\}_{j=1}^N$.

[10] Note that Bayes' Theorem states that

$$p(\Theta|Y^T, \mathcal{M}_i) = \frac{\mathcal{L}(\Theta|Y^T, \mathcal{M}_i)p(\Theta|\mathcal{M}_i)}{\int \mathcal{L}(\Theta|Y^T, \mathcal{M}_i)p(\Theta|\mathcal{M}_i)d\Theta}$$

But recognizing $\int \mathcal{L}(\Theta|Y^T, \mathcal{M}_i)p(\Theta|\mathcal{M}_i)d\Theta$ is constant for $\mathcal{M}_i$, we only need to be able to evaluate the posterior density up to a proportionate constant using

$$p(\Theta|Y^T, \mathcal{M}_i) \propto \mathcal{L}(\Theta|Y^T, \mathcal{M}_i)p(\Theta|\mathcal{M}_i)$$