Partial Adjustment As Optimal Response in a Dynamic Brainard Model

Richard Startz*

Revised January 2005

Abstract

Uncertainty about the precise quantitative effect of policy is endemic in economics. In a classic paper, Brainard showed that in the face of multiplier uncertainty in a static model that optimal policy is relatively conservative. I extend this work to a dynamic model and in the most simple case derive the classic partial adjustment model as the optimal response to shocks.

JEL Codes: C0, C1, E1

* University of Washington, Box 353330, Seattle, WA 98195 USA, email: startz@u.washington.edu.
The idea for this paper was suggested to me by Arabinda Basistha, who also made many suggestions while it was being constructed. Thoughtful comments from Stanley Fischer, Shelly Lundberg, Chang-Jin Kim, and Jeremy Piger are greatly appreciated.

It is a commonplace that in the face of uncertainty, policy should be applied cautiously. “Cautiously” is at times interpreted as meaning that policy should be applied gradually. By extending Brainard’s classic analysis to a dynamic setting we see in this paper that this is sometimes precisely the right advice.

Over 35 years ago William Brainard (Brainard, 1967) showed that when a policy multiplier is uncertain, one ought to aim to reach only part of the way toward the desired target, and that the optimal policy itself is applied more modestly than would be true under certainty equivalence. Operating in a static model Brainard cautioned “The gap [between optimal and certainty equivalence] in this context is not the difference between what policy was ‘last period’ and what would be required to make the expected value of [the target variable equal to the target.” (p. 415) Cautionary advice notwithstanding, Brainard’s work is frequently offered as an informal justification for gradual adjustment. In a dynamic model this can be justified rigorously.[1] Indeed, the primary contribution of this paper is to show that under a particular, reasonable specification, the optimal policy is to follow the classic partial adjustment model.[2]

Partial adjustment models have proven extraordinarily useful in empirical work and uncertainty as to the precise quantitative effect of manipulating a policy variable is endemic. While there is nothing in either Brainard’s analysis or the present one which limits its applicability to a specific branch of economics, both Brainard’s and recent work have been motivated by monetary policy concerns. Fischer and Cooper (1973) show that in a dynamic

[1] A point presaged perhaps by Kane’s commentary on the original presentation at the 1966 annual meetings, “As useful as this prospective should prove to be…it will be necessary to extend the Brainard model to dynamic situations.” (Kane 1967, p 432.)

[2] I start from a general model which gives a form of gradual adjustment, but in which classic partial adjustment is not the optimal solution. I then show the special assumptions needed for the classic partial adjustment model to apply exactly. The contrast may illustrate why partial adjustment is a useful approximation in some situations and not in others.

model with multiplier uncertainty, certainty equivalence policy is not optimal and that increased multiplier uncertainty argues for more cautious policy. (See also Cooper and Fischer, 1974.) Henderson and Turnovsky (1972) show that in a model in which quadratic adjustment costs for changes in the policy instrument lead to a partial adjustment model, increased multiplier uncertainty slows the rate of adjustment (although absent adjustment costs multiplier uncertainty does not generate partial adjustment.) Chow (1975) presents a general analysis of dynamic systems under uncertainty. Craine (1979) analyzes a problem very similar to the one presented below.

A number of recent papers have emphasized the theoretical and practical importance of gradual response in the context of interest rate smoothing by the Federal Reserve when implementing a modified Taylor rule, although these models do not develop the classic partial adjustment model. Clarida, Galí, and Gertler (2000) emphasizes the empirical importance of including a lagged interest rate in a monetary policy rule. Sack (2000) gives analytic results in a VAR context and uses numerical methods to provide empirical evidence that multiplier uncertainty matters considerably. Rudebusch (2001) applied numerical methods to a model of Fed interest rate smoothing, finding that uncertainty (at least as measured by estimated standard errors) is not very important. Wieland (2000, 2002) looks at parameter uncertainty that he then endogenizes, that is to say he then looks at the issue of learning and experimentation. See also Sack and Wieland (2000) and Svensson (1999).

I. The Static Brainard Model

I begin with a static Brainard model, both as a reminder of the classic result and to set out the basic mathematical structure of the model. In a static world one can write

$$y = \beta x + u \tag{1.1}$$

where y is the outcome, x is the policy instrument, β is the policy multiplier distributed $\left(\bar{\beta}, \sigma_\beta^2\right)$, and u is a shock distributed $\left(\bar{u}, \sigma_u^2\right)$, where both distributions are conditional on available information.

The objective function to be minimized is

$$L = E\left[\tfrac{1}{2}\left(y - y^*\right)^2\right] \tag{1.2}$$

One solves by setting the derivative w.r.t to the policy variable x equal to zero.

$$0 = \mathrm{E}\left[\left(\beta x + u - y^*\right)\beta\right] = x\,\mathrm{E}\left[\beta^2\right] + \mathrm{E}\left[u\beta\right] - y^*\,\mathrm{E}\left[\beta\right] \tag{1.3}$$

Two simplifications can be applied to equation (1.3). First, use the fact from statistics that $E\left[\beta^2\right] = \bar{\beta}^2 + \sigma_\beta^2$. Second, it is convenient and usually reasonable to assume that β and u are uncorrelated, in which case $\mathrm{E}\left[u\beta\right] = \mathrm{E}\left[u\right]\mathrm{E}\left[\beta\right]$. Using these simplifications the optimal policy is given in equation (1.4).

$$\begin{aligned} x &= \frac{\bar{\beta}^2}{\bar{\beta}^2 + \sigma_\beta^2} \cdot \frac{y^* - \bar{u}}{\bar{\beta}} \\ x &= \lambda \cdot \frac{y^* - \bar{u}}{\bar{\beta}}, \lambda \equiv \frac{\bar{\beta}^2}{\bar{\beta}^2 + \sigma_\beta^2} \end{aligned} \tag{1.4}$$

Equation (1.4) presents Brainard's classic result: optimal policy is a multiple λ, $0 \le \lambda \le 1$, of the certainty equivalence policy $\left(y^* - \bar{u}\right)/\bar{\beta}$. When there is no multiplier uncertainty, $\sigma_\beta^2 \to 0 \Rightarrow \lambda \to 1$, the certainty equivalence policy is optimal. The greater the uncertainty about the multiplier, scaled by its mean, the more "cautious" policy should be in the sense that the optimal policy moves toward zero.

Multiplier uncertainty arises for several reasons. Parameters are subject to estimation uncertainty, parameters evolve over time, and probably of greatest importance the "true" model is itself uncertain. One suspects that this last is the greatest source of uncertainty. But to illustrate that multiplier uncertainty can be an important practical issue, consider estimation uncertainty alone. Note that the ratio $\bar{\beta}/\sigma_{\beta}$ is "the *t*-statistic" from an econometric estimate. So that were estimation uncertainty the only issue, a *t*- of 2.0 would imply $\lambda = 0.8$ in equation (1.4).

In the static world described by equation (1.1) "cautious" does not imply any sort of gradual adjustment.[3] There is no temporal linkage between periods. If the effect of policy is very uncertain, then it is optimal to use very little policy in the sense that optimal x is close to zero. However, there is no sense in which one takes small steps. Equation (1.4) calls for a policy response which may be small, but which is complete in the current period. In the next section I turn to a model in which there are temporal linkages and where optimal policy does result in gradual adjustment.

II Dynamic Model Under Multiplier Uncertainty

In order for there to be persistence in policy response there needs to be persistence in the effect of policy. I make the model dynamic by allowing the *change* in y to be moved by both shocks and policy. In addition, I allow the policy multiplier to vary by period. In order to

[3] As Clarida, Galí, and Gertler (1999, page 1689) point out in reference to optimal interest rate rules, "…parameter uncertainty…may explain why … coefficients… are small relative to the case of certainty equivalence. But it does not explain…partial adjustment."

illustrate why specific assumptions are needed to derive the partial adjustment model, I start by solving a more general model in a finite horizon. I then give a formal derivation of the partial adjustment model with the required specialized assumptions by solving an infinite horizon dynamic program. The structural model is now

$$y_t = y_{t-1} + \beta_t x_t + u_t \tag{2.1}$$

where I assume that y_{t-1} is in the information set. The shock is composed of an anticipated part and a surprise, $u_t = {}_t\bar{u}_t + \tilde{u}_t$, where I use the prescript notation ${}_\tau\bar{u}_t$ to indicate the expectation of u_t formed at time τ. The variances of the two parts are $\sigma_{\bar{u}}^2$ and σ_u^2 respectively, and the anticipated shock and surprise are of course uncorrelated.

Let the intertemporal loss function be

$$\begin{aligned} \ell_t &= \frac{1}{2}\left(y_t - y^*\right)^2 \\ L_t &= \mathrm{E}\sum_{\tau=t}^{T} \rho^{\tau-t} \ell_\tau \end{aligned} \tag{2.2}$$

One wants to distinguish persistence due to multiplier uncertainty from the direct effect of the persistence in y due to building lagged y into the unit root specification in equation (2.1). Initially, consider the case where β_t is certain but u_t is random. Optimal policy is $x_t = \left(y^* - y_{t-1} - {}_t\bar{u}_t\right)/\beta_t$, which plugging back into equation (2.1) gives realized y_t.

$$y_t = y^* + u_t - {}_t\bar{u}_t \tag{2.3}$$

Lagging realized y_t one period and inserting back in the expression for x_t gives us the optimal policy under multiplier certainty

$$x_t^c = \frac{-1}{\beta_t}\left({}_t\overline{u}_t + \left(u_{t-1} - {}_{t-1}\overline{u}_{t-1}\right)\right) \tag{2.4}$$

Under multiplier certainty optimal policy responds fully to the expected part of the contemporaneous shock plus the unexpected part of the previous period's shock. The latter term in the optimal policy occurs because knowledge of y_{t-1} allows for correction of the previous period's error.

Note that $u_t - {}_t\overline{u}_t = \tilde{u}_t$ is an expectational error, and therefore serially uncorrelated, and more generally uncorrelated with any information available at the time the expectation was formed. Despite the unit root process in the structural equation, realized y_t, as given in equation (2.3), shows no persistence. Similarly, so long as ${}_t\overline{u}_t$ is serially uncorrelated optimal policy under multiplier certainty will be uncorrelated as well. So it should be clear that the unit root in the structural equation is not a propagation source of persistence.

Now introduce multiplier uncertainty into the dynamic model. Suppose that $\beta_t = \overline{\beta} + \varepsilon_t$ where $\varepsilon_t \sim \left(0, \sigma_\beta^2\right)$. I assume that the distribution of ε is ergodic and will assume shortly that the ε are i.i.d. I also make explicit the assumption that $y_{\tau-1}$ is in the information set iff $t > \tau - 1$ and assume that u and ε are independent at all leads and lags.

In period T the problem is

$$\min_{x_T} \mathrm{E}\,\tfrac{1}{2}\left(y_{T-1} + \beta_T x_T + u_T - y^*\right)^2 \tag{2.5}$$

Optimal period T policy is

$$x_T = \lambda \cdot \frac{y^* - y_{T-1} - {}_T\bar{u}_T}{\bar{\beta}}, \lambda \equiv \frac{\bar{\beta}^2}{\bar{\beta}^2 + \sigma_\beta^2} \tag{2.6}$$

It is useful to define the expected deviation of y from target in the absence of policy as $\tilde{y}_t \equiv y_{t-1} + {}_t\bar{u}_t - y^*$. One can then write the realized deviation of y from target as $y_t - y^* = \tilde{y}_t + \beta_t x_t + u_t - {}_t\bar{u}_t$.

Given the policy in equation (2.6) the deviation of realized y_T from target is

$$y_T - y^* = \tilde{y}_T \left(1 - \frac{\lambda}{\bar{\beta}} \beta_T\right) + \left(u_T - {}_T\bar{u}_T\right) \tag{2.7}$$

In period $T-1$ the decision-maker faces the problem $\min_{x_{T-1}} \mathrm{E}\left[\rho \ell_T + \ell_{T-1}\right]$, where the expectation operator $\mathrm{E}[\;]$ refers to the expectation taken at time $T-1$. The first partial of this objective function, using the independence of u and ε, is

$$\begin{aligned}
& x_{T-1}\, \mathrm{E}\left[\rho \beta_{T-1}^2 \left(1 - \frac{\lambda}{\bar{\beta}} \beta_T\right)^2 + \beta_{T-1}^2\right] + \\
& \mathrm{E}\left[y_{T-2} + u_{T-1} + {}_T\bar{u}_T - y^*\right] \cdot \rho \cdot \mathrm{E}\left[\beta_{T-1} \left(1 - \frac{\lambda}{\bar{\beta}} \beta_T\right)^2\right] + \\
& \mathrm{E}\left[u_T - {}_T\bar{u}_T\right] \cdot \rho \cdot \mathrm{E}\left[\beta_{T-1} \left(1 - \frac{\lambda}{\bar{\beta}} \beta_T\right)\right] + \\
& \mathrm{E}\left[y_{T-2} + u_{T-1} - y^*\right] \mathrm{E}\left[\beta_{T-1}\right]
\end{aligned} \tag{2.8}$$

A general solution to (2.8) involves high-order cross moments of ε. A paragraph hence I impose the assumption that the ε are i.i.d. Consider for a moment the more general case. The only situation in which there is not some generic form of gradual adjustment is if the response of x_{T-1} to $\mathrm{E}\left[y_{T-2} + u_{T-1} + {}_T\bar{u}_T - y^*\right]$ is $1/\beta$. As an example, suppose that β_T and β_{T-1} are perfectly correlated. The term multiplying x_{T-1} involves the second, third, and fourth central moments of β. The next term involves the first, second, and third central moments. The solution is untidy and does not equal $1/\beta$.[4] So the finding of lagged adjustment is quite general.[5]

The generic point about gradual adjustment doesn't depend on the assumption of i.i.d. errors, but the specific solution certainly does. Now assume that the ε are i.i.d.[6] In order to simplify equation (2.8) a few reminders about statistical algebra are helpful. First, by the law of iterated expectations $\mathrm{E}\left[u_T - {}_T\bar{u}_T\right] = {}_{T-1}\bar{u}_T - {}_{T-1}\bar{u}_T = 0$, eliminating the third term. Second, the expectation of the square of a random variable equals the square of the expectation plus the variance, so with a little algebra $\mathrm{E}\left[\left(1-\frac{\lambda}{\bar{\beta}}\beta_T\right)^2\right] = (1-\lambda)$. A third useful rule is that expectation of the product of independent random variables is the product of the expectations. For example, $\mathrm{E}\left[\beta_{T-1}^2\left(1-\frac{\lambda}{\bar{\beta}}\beta_T\right)^2\right] = \mathrm{E}\left[\beta_{T-1}^2\right]\cdot\mathrm{E}\left[\left(1-\frac{\lambda}{\bar{\beta}}\beta_T\right)^2\right]$ because ε_T and ε_{T-1} are independent. Using these three rules equation (2.8) simplifies to

$$\begin{aligned}
&x_{T-1}\left[\rho\left(\bar{\beta}^2+\sigma_\beta^2\right)(1-\lambda)+\left(\bar{\beta}^2+\sigma_\beta^2\right)\right]+\\
&\mathrm{E}\left[y_{T-2}+{}_{T-1}\bar{u}_{T-1}+{}_{T-1}\bar{u}_T-y^*\right]\cdot\rho\bar{\beta}(1-\lambda)+\\
&\mathrm{E}\left[y_{T-2}+{}_{T-1}\bar{u}_{T-1}-y^*\right]\bar{\beta}
\end{aligned} \qquad (2.9)$$

---

4 If one runs the problem back to period $T-2$ the analogous first-order condition runs to sixth moments, remains untidy, and still shows gradual adjustment, but the adjustment coefficients may be nonergodic – varying with time remaining to the terminal date.

5 However, remember that endogenous learning isn't being considered. If learning is important, a policy maker may engage in "excess adjustment" in order to produce better data with which to reduce future uncertainty. See Wieland (2000) for an example.

6 This is the same assumption made in Craine (1979). Quite clearly, there are situations in which independence is a good assumption and other situations in which it is not. In the latter case, uncertainty might not be a very good justification for assuming a simple partial adjustment mechanism.

Solving for optimal policy in period $T-1$ gives

$$x_{T-1} = \lambda \cdot \frac{y^* - y_{T-2} - {}_{T-1}\bar{u}_{T-1}}{\bar{\beta}} - \lambda \cdot \frac{{}_{T-1}\bar{u}_T}{\bar{\beta}} \cdot \frac{(1-\lambda)\rho}{(1-\lambda)\rho + 1} \tag{2.10}$$

According to equation (2.10), under uncertainty policy in period $T-1$ adjusts partially to shocks in period $T-1$. Specifically, the first term is λ times the certainty equivalence policy. In addition, policy partially anticipates shocks in period T as seen in the response to ${}_{T-1}\bar{u}_T$. But note that under certainty, $\lambda = 1$, policy does not anticipate future shocks. This is because the shock can be dealt with when it actually arrives. Also note that the discount rate matters only if the shocks are predictable a period in advance.[7]

The results for periods T and $T-1$ are generalized in the following lemma:

Lemma: If the shocks u_t are i.i.d.[8] then the policy rule is

$$x_t = \lambda \cdot \frac{y^* - y_{t-1} - {}_t\bar{u}_t}{\bar{\beta}} \tag{2.11}$$

Proof: Set this up as an infinite horizon stochastic dynamic programming problem. Define the optimal program

$$V(\tilde{y}_t) = \min_{x_t} \tfrac{1}{2} \mathrm{E}_t (y_t - y^*)^2 + \rho \mathrm{E}_t \left(V(\tilde{y}_{t+1}) \right) \tag{2.12}$$

As a prescient guess, suppose the value function can be written[9]

[7] The assumed independence of the ε and the quadratic form of the loss function also contribute to the discount rate dropping out of the solution.

[8] Note that the second term in equation (2.10) drops out because ${}_{T-1}\bar{u}_T = 0$.

[9] To verify that this is the proper value function, solve the optimization, insert equation (2.13) into equation (2.12) and show that the latter is indeed a valid equation. (Or see the appendix available from the author.)

$$V\left(\tilde{y}_t\right)=\frac{1}{2}\cdot\frac{1}{1-\rho(1-\lambda)}\left[(1-\lambda)\cdot\tilde{y}_t^2+\frac{1}{1-\rho}\cdot\sigma_u^2+\frac{\rho}{1-\rho}\cdot(1-\lambda)\cdot\sigma_{\bar{u}}^2\right] \tag{2.13}$$

In evaluating $V\left(\tilde{y}_{t+1}\right)$ make use of the substitution $\tilde{y}_{t+1}=\tilde{y}_t+\beta_t x_t+u_t-{}_t\bar{u}_t+{}_{t+1}\bar{u}_{t+1}$.

Noting that both $\partial y_t/\partial x_t=\beta_t$ and $\partial\tilde{y}_{t+1}/\partial x_t=\beta_t$, the first order condition for the dynamic program is

$$0=\mathrm{E}_t\left(\left(\tilde{y}_t+\beta_t x_t+u_t-{}_t\bar{u}_t\right)\beta_t\right)+\rho\,\mathrm{E}_t\left(\frac{1}{1-\rho(1-\lambda)}\cdot(1-\lambda)\cdot\tilde{y}_{t+1}\cdot\beta_t\right) \tag{2.14}$$

Substituting for $\tilde{y}_{t+1}$ one can re-write the first order condition as

$$0=\mathrm{E}_t\left(\left(\tilde{y}_t+\beta_t x_t+u_t-{}_t\bar{u}_t\right)\beta_t\left(1+\rho\frac{(1-\lambda)}{1-\rho(1-\lambda)}\right)+{}_{t+1}\bar{u}_{t+1}\cdot\beta_t\left(\rho\frac{(1-\lambda)}{1-\rho(1-\lambda)}\right)\right) \tag{2.15}$$

and, since ${}_t\bar{u}_{t+1}=0$ the optimal policy is

$$x_t=\lambda\frac{y^*-y_{t-1}-{}_t\bar{u}_t}{\bar{\beta}} \tag{2.16}$$

proving the lemma.

Deviations from target under optimal policy are

$$y_t-y^*=\tilde{y}_t\left[1-\beta_t\frac{\lambda}{\bar{\beta}}\right]+\left(u_t-{}_t\bar{u}_t\right) \tag{2.17}$$

and realized y obeys the process

$$y_t=(y_{t-1}+{}_t\bar{u}_t)\left[1-\beta_t\frac{\lambda}{\bar{\beta}}\right]+\left(u_t-{}_t\bar{u}_t\right)+\beta_t\frac{\lambda}{\bar{\beta}}y^* \tag{2.18}$$

In one sense the key to understanding why uncertainty leads to partial adjustment is seeing that the term in square brackets in equations (2.17) and (2.18) is nonzero. As a result, deviations from target this period are partially carried over to the future so that the policymaker next period is still responding to this period's shock. Contrast certainty where $\beta_t = \overline{\beta}$ and $\lambda = 1$ so the term in square brackets equals zero and y_t is affected only by the surprise to the contemporaneous shock.

III Impulse Response Functions and Partial Adjustment

Absent multiplier uncertainty, $\lambda = 1, \beta = \overline{\beta}$, optimal policy in equation (2.11) simplifies as $x_t^c = \left(y^* - y_{t-1} - {}_t\overline{u}_t \right) / \beta$. Similarly, absent multiplier uncertainty one can use $1 - \left(\lambda / \overline{\beta} \right) \beta_t = 0$ in equation (2.18) to find the realized value $y_t = y^* + u_t - {}_t\overline{u}_t$. Together these can be used to show that optimal policy responds fully to the expected part of the contemporaneous shock plus the unexpected part of the previous period's shock, $x_t^c = -\left(\left(u_{t-1} - {}_{t-1}\overline{u}_{t-1} \right) + {}_t\overline{u}_t \right) / \beta$, confirming the result derived directly in equation (2.4).

What is the impulse response function of x^c with respect to an anticipated negative unit shock? Contemporaneously x^c rises by $1/\beta$. The following period, both u_{t-1} and ${}_{t-1}\overline{u}_{t-1}$ have changed by one unit, so the effects offset. Thus the impulse response function of policy to an anticipated shock is a $1/\beta$ high spike followed by a zero flat as shown in Figure 1.

What is the impulse response function of x_t with respect to an anticipated negative unit shock when there is multiplier uncertainty? Contemporaneously x_t rises by $\lambda / \overline{\beta}$, from equation

(2.16). We can read the remainder of the impulse response from equations (2.16) and (2.18); the impulse response after N periods is

$$\frac{\lambda}{\bar{\beta}}\prod_{j=0}^{N-1}\left(1-\frac{\lambda}{\bar{\beta}}\beta_{t-N+j}\right) \tag{3.1}$$

The impulse response function in equation (3.1) is random, but we can usefully evaluate the path taking expectations with respect to β. The expected impulse response function to an anticipated shock, shown in Figure 1, is $\lambda/\bar{\beta}$, $\left(\lambda/\bar{\beta}\right)\cdot(1-\lambda)$, $\left(\lambda/\bar{\beta}\right)\cdot(1-\lambda)^2$, etc. The effect of policy on y cumulates over time. In the short run, the effect of policy is to offset in expected value a fraction λ of an anticipated shock. As the sum of the expected impulse response function is $1/\bar{\beta}$, in the long run an anticipated shock is fully offset.

So in the static Brainard model, adjustment is contemporaneous but conservative. In the dynamic model presented here, adjustment persists over time and in the long run there is full adjustment to the certainty equivalence level.

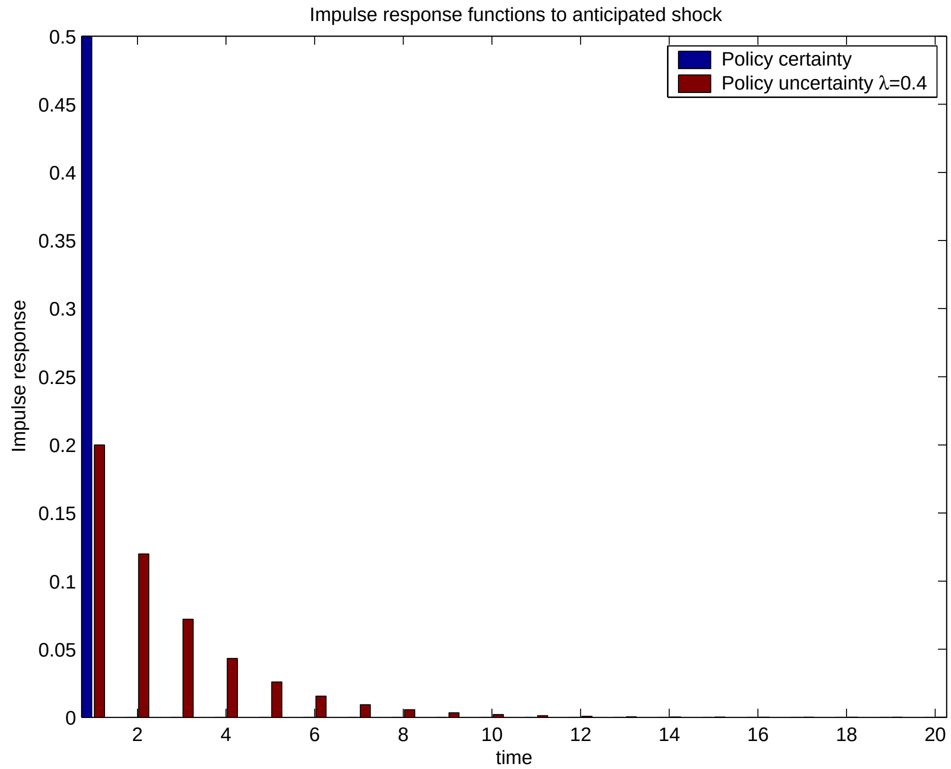

Figure 1

Turn now directly to the question of the classic partial adjustment model. Use repeated substitution on equation (2.17) to give

$$
\begin{aligned}
y_t - y^* = {}& \left(y_{t-N} - y^*\right)\prod_{j=0}^{N-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_{t-j}\right) \\
&+\sum_{n=0}^{N-1}\left(\left(u_{t-n} - \frac{\lambda}{\overline{\beta}}\beta_{t-n\ t-n}\overline{u}_{t-n}\right)\prod_{j=0}^{n-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_{t-n+1-j}\right)\right)
\end{aligned}
\tag{3.2}
$$

Suppose that at some point in the distant past the system was in equilibrium; label this period $t = 0$ so that $y_0 = y^*$ and the first term in equation (3.2) drops out. Insert a lagged version of equation (3.2) into equation (2.11) and write

$$x_t^* = -\frac{\lambda}{\overline{\beta}}\left[{}_t\overline{u}_t + \sum_{n=0}^{t-2}\left(\left(\left(u_{t-1-n} - {}_{t-1-n}\overline{u}_{t-1-n}\right) + {}_{t-1-n}\overline{u}_{t-1-n}\left(1-\frac{\lambda}{\overline{\beta}}\beta_{t-1-n}\right)\right)\prod_{j=0}^{n-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_{t-n-j}\right)\right)\right] \tag{3.3}$$

Since β is random x_t^* is random. Taking the time t expectation over β the expected path of x_t^* is

$$x_t^{*e} = -\frac{\lambda}{\overline{\beta}}\left[{}_t\overline{u}_t + \sum_{n=0}^{t-2}\left(\left[\left(u_{t-1-n} - {}_{t-1-n}\overline{u}_{t-1-n}\right) + {}_{t-1-n}\overline{u}_{t-1-n}\left(1-\lambda\right)\right]\left(1-\lambda\right)^{\max(0,n-1)}\right)\right] \tag{3.4}$$

where $\lim_{\lambda\to 1}(1-\lambda)^0 = 1$. Quasi-differencing gives

$$\begin{aligned} x_t^{*e} - (1-\lambda)x_{t-1}^{*e} &= -\frac{\lambda}{\overline{\beta}}\left({}_t\overline{u}_t + \left[u_{t-1} - {}_{t-1}\overline{u}_{t-1}\right] + {}_{t-1}\overline{u}_{t-1}(1-\lambda) - {}_{t-1}\overline{u}_{t-1}(1-\lambda)\right) \\ &= -\frac{\lambda}{\overline{\beta}}\left({}_t\overline{u}_t + \left[u_{t-1} - {}_{t-1}\overline{u}_{t-1}\right]\right) = \lambda x_t^c \end{aligned} \tag{3.5}$$

or

$$\mathrm{E}_t\left(x_t^*\right) - \mathrm{E}_{t-1}\left(x_{t-1}^*\right) = \lambda\left(x_t^c - \mathrm{E}_{t-1}\left(x_{t-1}^*\right)\right) \tag{3.6}$$

which is the classic partial adjustment model (Nerlove 1958).

IV Extensions

In this section I consider two extensions to the dynamic specification.

A. Serially correlated shocks without structural persistence

Suppose that the economy lacks structural persistence, $y_t = \beta x_t + u_t$, but that the shocks are serially correlated. If u_t is first order serially correlated, as in $u_t = \phi u_{t-1} + e_t$, then ${}_{t-1}\overline{u}_t = \phi\, {}_{t-1}\overline{u}_{t-1}$ and optimal policy is

$$x_{t-1} = \lambda \cdot \frac{y^* - {}_{t-1}\bar{u}_{t-1}}{\bar{\beta}}$$
$$x_t = \lambda \cdot \frac{y^* - \left(\phi_{t-1}\bar{u}_{t-1} + \left[{}_t\bar{u}_t - {}_{t-1}\bar{u}_t\right]\right)}{\bar{\beta}} \tag{4.1}$$

Noting that the term in square brackets in (4.1) is the surprise in the change in expectations between time $t-1$ and time t and therefore uncorrelated with information at time $t-1$, the correlation between policies in the two periods is $corr\left(x_t, x_{t-1}\right) = \phi$. The correlation is independent of λ. So while shocks are serially correlated and policy is both serially correlated and cautious, the serial correlation in policy has nothing to do with uncertainty and there is no gradual adjustment of policy.

B. Stationary dynamics

The static and dynamic models studied in the previous sections nest inside

$$y_t = \theta y_{t-1} + \beta_t x_t + u_t \tag{4.2}$$

with $\theta = 1$ or $\theta = 0$. Even when there is a unit root in the structural equation, $\theta = 1$, y_t is made stationary by the optimal policy. It is interesting to consider the case in which y_t is stationary because $\theta < 1$.

For general parameter values for λ and θ the optimal policy includes a nonergodic term in y^*. (See appendix available from the author.) Nonetheless, the expected value impulse response to an anticipated (negative) shock follows $\lambda/\bar{\beta}$, $\left[\lambda/\bar{\beta}\right]\cdot\left[\theta(1-\lambda)\right]$, $\left[\lambda/\bar{\beta}\right]\cdot\left[\theta(1-\lambda)\right]^2$, and so on. Except when dynamics are absent $(\theta = 0)$ or when uncertainty is absent $(\lambda = 1)$, the result of geometric decay of policy response to a shock is general, with the decay rate $\theta(1-\lambda)$ being more rapid when the persistence in y is less.

Note that the sum of the impulse response function is $\frac{1}{\bar{\beta}}\frac{\lambda}{1-\theta(1-\lambda)}$. The long-run response equals the certainty equivalent response only if $\theta = 1$.

V Conclusion

In a static framework multiplier uncertainty gives cause to reduce the magnitude of policy below the certainty equivalence level. In a dynamic framework multiplier uncertainty gives rise to gradual adjustment. With the stochastic specification used here, the classic partial adjustment model is optimal and in the long run there is full adjustment to the certainty equivalence level.

While the derivation of the classic partial adjustment model is tied to the model specification used here, the idea that gradual adjustment arises from a combination the presence of multiplier uncertainty and dynamics is quite general. If uncertainty leads to a conservative policy in period *t* so that shocks are not fully offset and the variable of interest is not brought fully to its target, and if the shortfall is propagated forward through time by the dynamics of the system, then in subsequent periods the policymaker will continue to respond to the gap. In Henderson and Turnovsky quadratic adjustment costs lead to partial adjustment – a propagation mechanism on to which multiplier uncertainty piggy-backs to further slow adjustment. In the model presented in section IVb above the autoregressive component built into the system provides the underlying propagation mechanism. Differing dynamic propagation methods and differing assumptions about the joint statistical distribution of multipliers over time lead to differing specifics of optimal policy, but the conclusion that the interaction of dynamics and uncertainty leads to gradual policy adjustment is quite general.

References

Brainard, William, “Uncertainty and the Effectiveness of Policy,” *American Economic Review*, May 1967, 57(2), pp. 411-425.

Chow, Gregory, *Analysis and Control of Dynamic Economic Systems*, John Wiley & Sons, 1975.

Clarida, Richard, Galí, Jordí, and Gertler, Mark, “The Science of Monetary Policy: A New Keynesian Perspective,” *Journal of Economic Literature*, Dec., 1999, pp. 1661-1707.

___________, “Monetary Policy Rules and Macroeconomic Stability,” *Quarterly-Journal-of-Economics,* 115(1), February 2000, pages 147-80.

Craine, Roger, “Optimal Monetary Policy with Uncertainty,” *Journal of Economic Dynamics and Control*, February 1979, 1(1), pp 59-83.

Cooper, J. Phillip and Fischer, Stanley, “Monetary and Fiscal Policy in the Fully Stochastic St. Louis Econometric Model,” *Journal of Money, Credit and Banking,* Vol. 6, No. 1, February 1974, pp. 1-22.

Fischer, Stanley and Cooper, J. Phillip, “Stabilization Policy and Lags,” *Journal of Political Economy,* Vol. 81, No. 4, Jul-Aug, 1973, pp. 847-877.

Henderson, Dale W. and Turnovsky, Stephen J., “Optimal Macroeconomic Policy Adjustment under Conditions of Risk,” *Journal of Economic Theory,* February 1972, Vol. 4, No. 1 pp. 58-71.

Kane, Edward J., “Discussion,” *American Economic Review*, May 1967,57(2), pp. 432-433.

Nerlove, Marc, *Distributed Lags and Demand Analysis for Agricultural and Other Commodities*, Washington, U.S. Department of Agriculture, Agricultural Handbook No. 141, June 1958.

Rudebusch, Glenn, "Is the Fed Too Timid? Monetary Policy in an Uncertain World," *Review of Economics and Statistics,* 83(2), May 2001, 203-217.

Sack, Brian, "Does the fed act gradually? A VAR analysis," *Journal of Monetary Economics*, August 2000, 46(1), pp 229-256.

Sack, Brian, and Wieland, Volcker, "Interest Rate Smoothing and Optimal Monetary Policy: A Review of Recent Empirical Evidence," *Journal of Economics and Business,* 2000, (52) pp. 205-228.

Svensson, Lars E.O., "Inflation Targeting: Some Extensions," *Scandinavian Journal of Economics,* 1999, 101(3), pp 337-361.

Wieland, Volker, "Monetary policy, parameter uncertainty and optimal learning," *Journal of Monetary Economics*, August 2000, 46(1), pp 199-228.

Wieland, Volker, "Monetary Policy and Uncertainty about the Natural Rate of Unemployment," Goethe University of Frankfurt, working paper 2002.

Appendix – Not For Publication

Appendix A

Suppose that rather than necessarily containing a unit root process, the specification for y_t is autoregressive as in

$$y_t = \theta y_{t-1} + \beta_t x_t + u_t, 0 \le \theta \le 1 \tag{A.1}$$

Equations (2.6) and (2.7) become

$$\begin{aligned} x_T &= \lambda \cdot \frac{y^* - \theta y_{T-1} - {}_T\bar{u}_T}{\bar{\beta}} \\ y_T - y^* &= \left(\theta y_{T-1} + {}_T\bar{u}_T - y^*\right)\left(1 - \frac{\lambda}{\bar{\beta}}\beta_T\right) + \left(u_T - {}_T\bar{u}_T\right) \end{aligned} \tag{A.2}$$

The problem in period $T-1$ becomes

$$\min_{x_{T-1}} \mathrm{E}\,\tfrac{1}{2}\left[\begin{aligned} &\rho\left(\left(\left(\theta\left(\theta y_{T-2} + \beta_{T-1}x_{T-1} + u_{T-1}\right)\right) + {}_T u_T - y^*\right)\left(1 - \frac{\lambda}{\bar{\beta}}\beta_T\right) + \left(u_T - {}_T\bar{u}_T\right)\right)^2 \\ &+\left(\theta y_{T-2} + \beta_{T-1}x_{T-1} + u_{T-1} - y^*\right)^2 \end{aligned}\right] \tag{A.3}$$

Taking first partials w.r.t. x_{T-1} we have

$$\mathrm{E}\left[\begin{aligned} &\rho\left(\left(\left(\theta\left(\theta y_{T-2} + \beta_{T-1}x_{T-1} + u_{T-1}\right)\right) + {}_T\bar{u}_T - y^*\right)\left(1 - \frac{\lambda}{\bar{\beta}}\beta_T\right) + \left(u_T - {}_T\bar{u}_T\right)\right)\left(\theta\beta_{T-1}\left(1 - \frac{\lambda}{\bar{\beta}}\beta_T\right)\right) \\ &+\left(\theta y_{T-2} + \beta_{T-1}x_{T-1} + u_{T-1} - y^*\right)\left(\beta_{T-1}\right) \end{aligned}\right] \tag{A.4}$$

Re-arranging (A.4) and using the independence of u and ε gives

$$
\begin{aligned}
& x_{T-1}\,\mathrm{E}\left\{\rho\theta^2\beta^2_{T-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)^2+\beta^2_{T-1}\right\} \\
& +\theta y_{T-2}\,\mathrm{E}\left[\rho\theta^2\beta_{T-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)^2+\beta_{T-1}\right] \\
& +\mathrm{E}\left[u_T-{}_T\overline{u}_T\right]\mathrm{E}\left\{\rho\theta\beta_{T-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)\right\} \\
& +\mathrm{E}\left[u_{T-1}\right]\mathrm{E}\left\{\rho\theta^2\beta_{T-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)^2+\beta_{T-1}\right\} \\
& +\mathrm{E}\left[{}_T\overline{u}_T\right]\mathrm{E}\left\{\rho\theta\beta_{T-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)^2\right\} \\
& -y^*\,\mathrm{E}\left\{\rho\theta\beta_{T-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)^2+\beta_{T-1}\right\}
\end{aligned}
\tag{A.5}
$$

By the law of iterated expectations $\mathrm{E}\left[u_T-{}_T\overline{u}_T\right]={}_{T-1}\overline{u}_T-{}_{T-1}\overline{u}_T=0$. Further simplifications arise from the facts that $\mathrm{E}\left[\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)^2\right]=\left(1-\lambda\right)$ and $\mathrm{E}\left[\beta^2_{T-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)^2\right]=\mathrm{E}\left[\beta^2_{T-1}\right]\cdot\mathrm{E}\left[\left(1-\frac{\lambda}{\overline{\beta}}\beta_T\right)^2\right]=\left(\overline{\beta}^2+\sigma^2_\beta\right)\left(1-\lambda\right)$. We can re-write (A.5) as

$$
\begin{aligned}
& x_{T-1}\left\{\left(\rho\theta^2\left(1-\lambda\right)+1\right)\left(\overline{\beta}^2+\sigma^2_\beta\right)\right\} \\
& +\left[\theta y_{T-2}+{}_{T-1}\overline{u}_{T-1}\right]\left\{\left(\rho\theta^2\left(1-\lambda\right)+1\right)\overline{\beta}\right\} \\
& +{}_{T-1}\overline{u}_T\left\{\rho\theta\overline{\beta}\left(1-\lambda\right)\right\} \\
& -y^*\left\{\rho\theta\overline{\beta}\left(1-\lambda\right)+\overline{\beta}\right\}
\end{aligned}
\tag{A.6}
$$

Setting (A.6) equal to zero and assuming ${}_{T-1}\overline{u}_T=0$ gives

$$x_{T-1} = -\lambda \frac{\theta y_{T-2} + {}_{T-1}\bar{u}_{T-1}}{\bar{\beta}} + \lambda \frac{y^*}{\bar{\beta}} \cdot \frac{\rho\theta(1-\lambda)+1}{\rho\theta^2(1-\lambda)+1}$$
$$= -\frac{\lambda}{\bar{\beta}}\left[\theta y_{T-2} + {}_{T-1}\bar{u}_{T-1} - y^* \cdot \frac{\rho\theta(1-\lambda)+1}{\rho\theta^2(1-\lambda)+1}\right] \tag{A.7}$$

Note that equation (A.7) nests the two special cases given in the text: $\theta = 1$ and $\theta = 0$. Unlike the solution given in (2.11) the formula for the optimal policy is not generally ergodic because of the term multiplying y^*. (Although the policy rule is ergodic if any of $\theta = 0$, $\theta = 1$, or $\lambda = 1$ are true. Nonetheless, the expected value impulse response to an anticipated shock follows $-\lambda/\bar{\beta}$, $\left[-\lambda/\bar{\beta}\right]\cdot\left[\theta(1-\lambda)\right]$, $\left[-\lambda/\bar{\beta}\right]\cdot\left[\theta(1-\lambda)\right]^2$, and so on. Thus the result of geometric decay is general, except when dynamics are absent $(\theta = 0)$ or when uncertainty is absent $(\lambda = 1)$.

As a check, run the problem back one more period. Using (A.7) we have

$$y_{T-1} - y^* = \theta y_{T-2} + u_{T-1} - y^* + \beta_{T-1} x_{T-1}$$
$$= \theta y_{T-2} + u_{T-1} - y^* + \beta_{T-1}\left(-\frac{\lambda}{\bar{\beta}}\left[\theta y_{T-2} + {}_{T-1}\bar{u}_{T-1} - y^* \cdot \frac{\rho\theta(1-\lambda)+1}{\rho\theta^2(1-\lambda)+1}\right]\right)$$
$$= \theta\left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1}\right) y_{T-2} - y^*\left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1} \cdot \frac{\rho\theta(1-\lambda)+1}{\rho\theta^2(1-\lambda)+1}\right) \tag{A.8}$$
$$+\left(u_{T-1} - {}_{T-1}\bar{u}_{T-1}\right) + \left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1}\right){}_{T-1}\bar{u}_{T-1}$$

Using (A.2) we have

$$y_T - y^* = \left(\theta\left(y_{T-1}-y^*\right) + {}_T\bar{u}_T - (1-\theta)\,y^*\right)\left(1-\frac{\lambda}{\bar{\beta}}\beta_T\right) + \left(u_T - {}_T\bar{u}_T\right)$$

$$= \left(\begin{array}{l} \theta\left\{\begin{array}{l} \theta\left(1-\frac{\lambda}{\bar\beta}\beta_{T-1}\right)y_{T-2} - y^*\left(1-\frac{\lambda}{\bar\beta}\beta_{T-1}\cdot\frac{\rho\theta(1-\lambda)+1}{\rho\theta^2(1-\lambda)+1}\right) \\ +\left(u_{T-1}-{}_{T-1}\bar u_{T-1}\right)+\left(1-\frac{\lambda}{\bar\beta}\beta_{T-1}\right){}_{T-1}\bar u_{T-1}\end{array}\right\} \\ + {}_T\bar u_T - (1-\theta)\,y^*\end{array}\right)\left(1-\frac{\lambda}{\bar\beta}\beta_T\right)+\left(u_T-{}_T\bar u_T\right) \tag{A.9}$$

The problem in period $T-2$ becomes

$$\min_{x_{T-2}} \mathrm{E}\,\tfrac{1}{2}\left[\rho^2\left(y_T-y^*\right)^2+\rho\left(y_{T-1}-y^*\right)^2+\left(y_{T-2}-y^*\right)^2\right]$$

which has the first-order condition

$$0=\mathrm{E}\left[\left(\rho^2\left(y_T-y^*\right)\cdot\frac{\partial y_T}{\partial x_{T-2}}+\rho\left(y_{T-1}-y^*\right)\cdot\frac{\partial y_{T-1}}{\partial x_{T-2}}+\left(y_{T-2}-y^*\right)\cdot\frac{\partial y_{T-2}}{\partial x_{T-2}}\right)\right] \tag{A.10}$$

It is useful to note that for any fixed ω, $\mathrm{E}\left[\left(1-\frac{\lambda}{\bar\beta}\beta_t\right)\left(1-\frac{\lambda}{\bar\beta}\beta_t\omega\right)\right]=1-\lambda$.

Taking expectations term by term gives for the first term

$$\mathrm{E}\left[\left(\rho^2\left(y_T-y^*\right)\cdot\frac{\partial y_T}{\partial x_{T-2}}\right)\right]=$$

$$\mathrm{E}\left[\begin{array}{l}\rho^2\left(\left(\begin{array}{l}\theta\left\{\begin{array}{l}\theta\left(1-\frac{\lambda}{\bar\beta}\beta_{T-1}\right)\left(\theta y_{T-3}+\beta_{T-2}x_{T-2}+u_{T-2}\right)-y^*\left(1-\frac{\lambda}{\bar\beta}\beta_{T-1}\cdot\frac{\rho\theta(1-\lambda)+1}{\rho\theta^2(1-\lambda)+1}\right)\\ +\left(u_{T-1}-{}_{T-1}\bar u_{T-1}\right)+\left(1-\frac{\lambda}{\bar\beta}\beta_{T-1}\right){}_{T-1}\bar u_{T-1}\end{array}\right\}\\ +{}_T\bar u_T-(1-\theta)\,y^*\end{array}\right)\right)\\ \cdot\left(\theta^2\left(1-\frac{\lambda}{\bar\beta}\beta_{T-1}\right)\beta_{T-2}\right)\end{array}\right] \tag{A.11}$$

which simplifies to

$$
\begin{aligned}
&= \rho^2 \begin{bmatrix} \left(\theta y_{T-3} + {}_{T-2}\bar{u}_{T-2}\right)\left(\theta^4 (1-\lambda)\bar{\beta}\right) + x_{T-2}\left(\theta^4 (1-\lambda)\left(\bar{\beta}^2 + \sigma^2\right)\right) \\ -y^* \theta^2 \left(\theta(1-\lambda)\bar{\beta} + (1-\theta)(1-\lambda)\bar{\beta}\right) \end{bmatrix} \\
&= \rho^2 \begin{bmatrix} \theta^4 (1-\lambda)\left(\bar{\beta}\left(\theta y_{T-3} + {}_{T-2}\bar{u}_{T-2}\right) + \left(\bar{\beta}^2 + \sigma^2\right) x_{T-2}\right) \\ -y^* \theta^2 (1-\lambda)\bar{\beta} \end{bmatrix}
\end{aligned}
\tag{A.12}
$$

Taking expectation of the second term gives

$$
\begin{aligned}
&\mathrm{E}\left[\rho\left(y_{T-1} - y^*\right) \cdot \frac{\partial y_{T-1}}{\partial x_{T-2}}\right] = \\
&\mathrm{E}\begin{bmatrix} \rho \begin{pmatrix} \theta\left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1}\right)\left(\theta y_{T-3} + \beta_{T-2} x_{T-2} + u_{T-2}\right) - y^*\left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1} \cdot \frac{\rho\theta(1-\lambda)+1}{\rho\theta^2(1-\lambda)+1}\right) \\ + \left(u_{T-1} - {}_{T-1}\bar{u}_{T-1}\right) + \left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1}\right) {}_{T-1}\bar{u}_{T-1} \end{pmatrix} \\ \cdot \left(\theta\left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1}\right)\beta_{T-2}\right) \end{bmatrix}
\end{aligned}
\tag{A.13}
$$

which simplifies to

$$
\begin{aligned}
&= \rho \begin{bmatrix} \begin{pmatrix} \theta\left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1}\right)\left(\theta y_{T-3} + \beta_{T-2} x_{T-2} + {}_{T-2}\bar{u}_{T-2}\right) - y^*\left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1} \cdot \frac{\rho\theta(1-\lambda)+1}{\rho\theta^2(1-\lambda)+1}\right) \\ + \left(u_{T-1} - {}_{T-1}\bar{u}_{T-1}\right) + \left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1}\right) {}_{T-1}\bar{u}_{T-1} \end{pmatrix} \\ \cdot \left(\theta\left(1 - \frac{\lambda}{\bar{\beta}}\beta_{T-1}\right)\beta_{T-2}\right) \end{bmatrix} \\
&= \rho \begin{bmatrix} \theta^2 (1-\lambda)\left(\bar{\beta}\left(\theta y_{T-3} + {}_{T-2}\bar{u}_{T-2}\right) + \left(\bar{\beta}^2 + \sigma^2\right) x_{T-2}\right) \\ -y^* \left(\theta(1-\lambda)\bar{\beta}\right) \end{bmatrix}
\end{aligned}
\tag{A.14}
$$

Taking expectation of the third term gives

$$
\begin{aligned}
&\mathrm{E}\left[\left(y_{T-2}-y^{*}\right)\cdot\frac{\partial y_{T-2}}{\partial x_{T-2}}\right]= \\
&\mathrm{E}\left[\left(\left(\theta y_{T-3}+\beta_{T-2}x_{T-2}+u_{T-2}\right)-y^{*}\right)\cdot\beta_{T-2}\right]
\end{aligned}
\tag{A.15}
$$

which simplifies to

$$
\begin{aligned}
&\mathrm{E}\left[\left(\left(\theta y_{T-3}+\beta_{T-2}x_{T-2}+u_{T-2}\right)-y^{*}\right)\cdot\beta_{T-2}\right] \\
&=\left(\theta y_{T-3}+{}_{T-2}\bar{u}_{T-2}-y^{*}\right)\bar{\beta}+\left(\bar{\beta}^{2}+\sigma^{2}\right)x_{T-2}
\end{aligned}
\tag{A.16}
$$

Putting the first order condition back together gives

$$
\begin{aligned}
0=\rho^{2}&\begin{bmatrix}\theta^{4}(1-\lambda)\left(\bar{\beta}\left(\theta y_{T-3}+{}_{T-2}\bar{u}_{T-2}\right)+\left(\bar{\beta}^{2}+\sigma^{2}\right)x_{T-2}\right)\\ -y^{*}\theta^{2}(1-\lambda)\bar{\beta}\end{bmatrix}\\
+\rho&\begin{bmatrix}\theta^{2}(1-\lambda)\left(\bar{\beta}\left(\theta y_{T-3}+{}_{T-2}\bar{u}_{T-2}\right)+\left(\bar{\beta}^{2}+\sigma^{2}\right)x_{T-2}\right)\\ -y^{*}\left(\theta(1-\lambda)\bar{\beta}\right)\end{bmatrix}\\
&+\left(\theta y_{T-3}+u_{T-2}-y^{*}\right)\bar{\beta}+\left(\bar{\beta}^{2}+\sigma^{2}\right)x_{T-2}
\end{aligned}
\tag{A.17}
$$

Collecting terms gives us

$$
\begin{aligned}
0=&\left(\theta y_{T-3}+{}_{T-2}\bar{u}_{T-2}\right)\left(\rho^{2}\theta^{4}(1-\lambda)\bar{\beta}+\rho\theta^{2}(1-\lambda)\bar{\beta}+\bar{\beta}\right)\\
&+x_{T-2}\left(\rho^{2}\theta^{4}(1-\lambda)\left(\bar{\beta}^{2}+\sigma^{2}\right)+\rho\theta^{2}(1-\lambda)\left(\bar{\beta}^{2}+\sigma^{2}\right)+\left(\bar{\beta}^{2}+\sigma^{2}\right)\right)\\
&-y^{*}\left(\rho^{2}\left(\theta(1-\lambda)\bar{\beta}\right)+\rho\left(\theta(1-\lambda)\bar{\beta}\right)+\bar{\beta}\right)\\
0=&\left(\theta y_{T-3}+{}_{T-2}\bar{u}_{T-2}\right)\bar{\beta}\left(\rho^{2}\theta^{4}(1-\lambda)+\rho\theta^{2}(1-\lambda)+1\right)\\
&+x_{T-2}\left(\bar{\beta}^{2}+\sigma^{2}\right)\left(\rho^{2}\theta^{4}(1-\lambda)+\rho\theta^{2}(1-\lambda)+1\right)\\
&-y^{*}\bar{\beta}\left(\rho^{2}\left(\theta(1-\lambda)\right)+\rho\left(\theta(1-\lambda)\right)+1\right)
\end{aligned}
\tag{A.18}
$$

Optimal policy is given by

$$
\begin{aligned}
0 = &\left(\theta y_{T-3} + {}_{T-2}\bar{u}_{T-2}\right)\bar{\beta}\left(\rho^2\theta^4(1-\lambda)+\rho\theta^2(1-\lambda)+1\right)\\
&+x_{T-2}\left(\bar{\beta}^2+\sigma^2\right)\left(\rho^2\theta^4(1-\lambda)+\rho\theta^2(1-\lambda)+1\right)\\
&-y^*\bar{\beta}\left(\rho^2\left(\theta(1-\lambda)\right)+\rho\left(\theta(1-\lambda)\right)+1\right)
\end{aligned}
\tag{A.19}
$$

$$
x_{T-2} = -\frac{\lambda}{\bar{\beta}}\left(\theta y_{T-3} + {}_{T-2}\bar{u}_{T-2}\right) + \frac{\lambda}{\bar{\beta}} y^* \left(\frac{\rho^2\theta(1-\lambda)+\rho\theta(1-\lambda)+1}{\rho^2\theta^4(1-\lambda)+\rho\theta^2(1-\lambda)+1} \right)
$$

It appears that the general solution for optimal policy is

$$
x_{T-\tau} = -\frac{\lambda}{\bar{\beta}}\left(\theta y_{T-\tau-1} + {}_{T-\tau}\bar{u}_{T-\tau}\right) + \frac{\lambda}{\bar{\beta}} y^* \left(\frac{1+(1-\lambda)\theta\left(0+\rho+\rho^2+\cdots+\rho^\tau\right)}{1+(1-\lambda)\left(0+\left(\rho\theta^2\right)+\left(\rho\theta^2\right)^2+\cdots+\left(\rho\theta^2\right)^\tau\right)} \right) \tag{A.20}
$$

Note that if any of $\theta = 1$, $\theta = 0$, or $\lambda = 1$ are true, then the last factor in equation (A.20) equals one and the formula reduces to that given earlier in the paper.

Appendix B

Here is an alternative derivation of the first lemma in the paper.

Lemma: If the shocks u_t are i.i.d.[10] then the policy rule is

$$x_t = \lambda \cdot \frac{y^* - y_{t-1} - {}_t\overline{u}_t}{\overline{\beta}} \tag{B.1}$$

Proof: By induction on $N = 0 \ldots T$ for $t = T - N$. Equations (2.6) and (2.10) give the proof for $N = 0,1$. Assume the lemma for $N-1$ and prove for N.

The general formula for deviations of realized y_t from target is

$$y_t - y^* = \left(y_{t-1} - y^*\right)\left(1 - \frac{\lambda}{\overline{\beta}}\beta_t\right) + u_t - \frac{\lambda}{\overline{\beta}}\beta_t\,{}_t\overline{u}_t \tag{B.2}$$

By repeated substitution (B.2) can be written

$$\begin{aligned} y_t - y^* &= \left(y_{t-N} - y^*\right)\prod_{j=0}^{N-1}\left(1 - \frac{\lambda}{\overline{\beta}}\beta_{t-j}\right) \\ &\quad + \sum_{n=0}^{N-1}\left(\left(u_{t-n} - \frac{\lambda}{\overline{\beta}}\beta_{t-n}\,{}_{t-n}\overline{u}_{t-n}\right)\prod_{j=0}^{n-1}\left(1 - \frac{\lambda}{\overline{\beta}}\beta_{t-n+1-j}\right)\right) \end{aligned} \tag{B.3}$$

where as a convention $\prod_{\tau=t}^{k} f(\tau) = 1$ if $k < t$.

[10] Note that the second term in equation (2.10) drops out because ${}_{T-1}\overline{u}_T = 0$.

To minimize L_{T-N} we need the derivative of ℓ_τ w.r.t. $x_{\tau-N}$ which is

$\mathrm{E}\left[\left(y_\tau - y^*\right)\beta_{\tau-N}\prod_{j=0}^{N-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_{\tau-j}\right)\right]$, where the expectation is taken at time $T-N$. Because u_t is i.i.d. ${}_{T-N}\overline{u}_{T-n} = 0, n < N$, so the second term in (B.3) drops out. We can write

$$\begin{aligned}\frac{\partial \mathrm{E}\,\ell_\tau}{\partial x_{\tau-N}} &= \mathrm{E}\left[\left(y_{\tau-N-1} + \beta_{\tau-N}x_{\tau-N} + u_{\tau-N} - y^*\right)\beta_{\tau-N}\prod_{j=0}^{N-1}\left(1-\frac{\lambda}{\overline{\beta}}\beta_{\tau-j}\right)\right] \\ &= x_{\tau-N}\left(\overline{\beta}^2 + \sigma_\beta^2\right)\left(1-\lambda\right)^N + \left(y_{\tau-N-1} + {}_{\tau-N}\overline{u}_{\tau-N} - y^*\right)\overline{\beta}\left(1-\lambda\right)^N\end{aligned} \tag{B.4}$$

Now take the derivative of L_t w.r.t. x_t.

$$\frac{\partial L_t}{\partial x_t} = \sum_{\tau=t}^{T}\rho^{\tau-t}\frac{\partial \mathrm{E}\,\ell_\tau}{\partial x_t} = \sum_{\tau=t}^{T}\rho^{\tau-t}\left\{x_t\left(\overline{\beta}^2 + \sigma_\beta^2\right)\left(1-\lambda\right)^{\tau-t} + \left(y_{t-1} + {}_t\overline{u}_t - y^*\right)\overline{\beta}\left(1-\lambda\right)^{\tau-t}\right\} \tag{B.5}$$

Setting the expression in (B.5) equal to zero and noting that variables not involving τ pass through the summation operator, the optimization problem solves

$$0 = x_t\left(\overline{\beta}^2 + \sigma_\beta^2\right)\sum_{\tau=t}^{T}\rho^{\tau-t}\left(1-\lambda\right)^{\tau-t} + \left(y_{t-1} + {}_t\overline{u}_t - y^*\right)\overline{\beta}\sum_{\tau=t}^{T}\rho^{\tau-t}\left(1-\lambda\right)^{\tau-t} \tag{B.6}$$

The summation terms in equation (B.6) can be cancelled out. What's left gives the optimal policy as specified in equation (B.1), completing the proof.

Appendix C

The following is the proof that the proposed value function "works."

The value function is

$$V\left(\tilde{y}_t\right)=\frac{1}{2}\cdot\frac{1}{1-\rho\left(1-\lambda\right)}\left[\left(1-\lambda\right)\cdot\tilde{y}_t^2+\frac{1}{1-\rho}\cdot\sigma_u^2+\frac{\rho}{1-\rho}\cdot\left(1-\lambda\right)\cdot\sigma_{\bar{u}}^2\right] \tag{C.1}$$

We need to show that

$$V\left(\tilde{y}_t\right)=\min_{x_t}\tfrac{1}{2}\mathrm{E}_t\left(y_t-y*\right)^2+\rho\,\mathrm{E}_t\left(V\left(\tilde{y}_{t+1}\right)\right) \tag{C.2}$$

is a valid equation. Two useful substitutions are

$$\begin{aligned}
y_t-y^* &= \tilde{y}_t+\beta_t x_t+u_t-{}_t\bar{u}_t\\
&= \tilde{y}_t+\beta_t\left(\lambda\frac{y^*-y_{t-1}-{}_t\bar{u}_t}{\bar{\beta}}\right)+u_t-{}_t\bar{u}_t\\
&= \tilde{y}_t\left[1-\beta_t\frac{\lambda}{\bar{\beta}}\right]+\left(u_t-{}_t\bar{u}_t\right)\\
\tilde{y}_{t+1} &= y_t+{}_{t+1}\bar{u}_{t+1}-y^*\\
&= \tilde{y}_t\left[1-\beta_t\frac{\lambda}{\bar{\beta}}\right]+\left(u_t-{}_t\bar{u}_t\right)+{}_{t+1}\bar{u}_{t+1}
\end{aligned} \tag{C.3}$$

Evaluating the forward term gives

$$
\begin{aligned}
\mathrm{E}_t\left(V\left(\tilde{y}_{t+1}\right)\right) &= \mathrm{E}_t\left(V\left(\tilde{y}_t\left[1-\beta_t\frac{\lambda}{\overline{\overline{\beta}}}\right]+\left(u_t-{}_t\overline{u}_t\right)+{}_{t+1}\overline{u}_{t+1}\right)\right) \\
&= \frac{1}{2}\cdot\frac{1}{1-\rho\left(1-\lambda\right)}\left[\left(1-\lambda\right)\cdot\mathrm{E}_t\left(\tilde{y}_t\left[1-\beta_t\frac{\lambda}{\overline{\overline{\beta}}}\right]+\left(u_t-{}_t\overline{u}_t\right)+{}_{t+1}\overline{u}_{t+1}\right)^2+\frac{1}{1-\rho}\cdot\sigma_u^2+\frac{\rho}{1-\rho}\cdot\left(1-\lambda\right)\cdot\sigma_{\overline{u}}^2\right] \\
&= \frac{1}{2}\cdot\frac{1}{1-\rho\left(1-\lambda\right)}\left[\left(1-\lambda\right)\cdot\left(\tilde{y}_t^2\left(1-\lambda\right)+\sigma_u^2+\sigma_{\overline{u}}^2\right)+\frac{1}{1-\rho}\cdot\sigma_u^2+\frac{\rho}{1-\rho}\cdot\left(1-\lambda\right)\cdot\sigma_{\overline{u}}^2\right]
\end{aligned}
\tag{C.4}
$$

The penalty function is

$$
\begin{aligned}
\mathrm{E}\left(y_t-y^*\right)^2 &= \mathrm{E}\left(\tilde{y}_t\left[1-\beta_t\frac{\lambda}{\overline{\overline{\beta}}}\right]+\left(u_t-{}_t\overline{u}_t\right)\right)^2 \\
&= \tilde{y}_t^2\left(1-\lambda\right)+\sigma_u^2
\end{aligned}
\tag{C.5}
$$

Substituting into the original equation gives

$$
\begin{aligned}
V\left(\tilde{y}_t\right) &= \min_{x_t}\tfrac{1}{2}\mathrm{E}_t\left(y_t-y*\right)^2+\rho\,\mathrm{E}_t\left(V\left(\tilde{y}_{t+1}\right)\right) \\
&\frac{1}{2}\cdot\frac{1}{1-\rho\left(1-\lambda\right)}\left[\left(1-\lambda\right)\cdot\tilde{y}_t^2+\frac{1}{1-\rho}\cdot\sigma_u^2+\frac{\rho}{1-\rho}\cdot\left(1-\lambda\right)\cdot\sigma_{\overline{u}}^2\right] \\
&= \tfrac{1}{2}\left(\tilde{y}_t^2\left(1-\lambda\right)+\sigma_u^2\right) \\
&+\rho\left[\frac{1}{2}\cdot\frac{1}{1-\rho\left(1-\lambda\right)}\left[\left(1-\lambda\right)\cdot\left(\tilde{y}_t^2\left(1-\lambda\right)+\sigma_u^2+\sigma_{\overline{u}}^2\right)+\frac{1}{1-\rho}\cdot\sigma_u^2+\frac{\rho}{1-\rho}\cdot\left(1-\lambda\right)\cdot\sigma_{\overline{u}}^2\right]\right]
\end{aligned}
\tag{C.6}
$$

multiply through by $2\left(1-\rho\left(1-\lambda\right)\right)$ and collect terms giving

$$
\begin{aligned}
&\left[\left(1-\lambda\right)\cdot\tilde{y}_t^2+\frac{1}{1-\rho}\cdot\sigma_u^2+\frac{\rho}{1-\rho}\cdot\left(1-\lambda\right)\cdot\sigma_{\overline{u}}^2\right] \\
&=\left(1-\rho\left(1-\lambda\right)\right)\left(\tilde{y}_t^2\left(1-\lambda\right)+\sigma_u^2\right) \\
&+\rho\left[\left[\left(1-\lambda\right)\cdot\left(\tilde{y}_t^2\left(1-\lambda\right)+\sigma_u^2+\sigma_{\overline{u}}^2\right)+\frac{1}{1-\rho}\cdot\sigma_u^2+\frac{\rho}{1-\rho}\cdot\left(1-\lambda\right)\cdot\sigma_{\overline{u}}^2\right]\right]
\end{aligned}
\tag{C.7}
$$

Now collect terms as in

$$\begin{aligned}&\left[(1-\lambda)-\left(1-\rho(1-\lambda)\right)(1-\lambda)-\rho(1-\lambda)^2\right]\cdot\tilde{y}_t^2\\&+\left[\frac{1}{1-\rho}-\left(1-\rho(1-\lambda)\right)-\rho(1-\lambda)-\frac{\rho}{1-\rho}\right]\cdot\sigma_u^2\\&+\left[\frac{\rho}{1-\rho}\cdot(1-\lambda)-\rho(1-\lambda)-\frac{\rho^2}{1-\rho}\cdot(1-\lambda)\right]\cdot\sigma_{\bar{u}}^2\end{aligned}\tag{C.8}$$

Each of the three leading coefficients equals zero, proving the equation is valid.

Appendix D

Miscellaneous bits of math that may prove useful to the reader are collected here.

In section 2, it may help to write out the problem at time $T-1$ as

$$\min_{x_{T-1}} \mathrm{E}\,\tfrac{1}{2}\left[\begin{array}{l}\rho\left(\left(\left(\left(y_{T-2}+\beta_{T-1}x_{T-1}+u_{T-1}\right)\right)+{}_{T}\bar{u}_{T}-y^{*}\right)\left(1-\dfrac{\lambda}{\bar{\beta}}\beta_{T}\right)+\left(u_{T}-{}_{T}\bar{u}_{T}\right)\right)^{2}\\ +\left(y_{T-2}+\beta_{T-1}x_{T-1}+u_{T-1}-y^{*}\right)^{2}\end{array}\right] \tag{D.1}$$

The proof that $\mathrm{E}\left[\left(1-\frac{\lambda}{\bar{\beta}}\beta_{T}\right)^{2}\right]=\left(1-\lambda\right)$ comes from

$$\begin{aligned}\mathrm{E}\left[\left(1-\frac{\lambda}{\bar{\beta}}\beta_{T}\right)^{2}\right]&=\left(1-\frac{\lambda}{\bar{\beta}}\bar{\beta}\right)^{2}+\left(\frac{\lambda}{\bar{\beta}}\right)^{2}\sigma_{\beta}^{2}\\&=\left(1-\lambda\right)^{2}+\lambda\left(\frac{\bar{\beta}^{2}}{\bar{\beta}^{2}+\sigma_{\beta}^{2}}\right)\left(\frac{1}{\bar{\beta}}\right)^{2}\sigma_{\beta}^{2}\\&=\left(1-\lambda\right)^{2}+\lambda\left(1-\lambda\right)\\&=\left(1-\lambda\right)\end{aligned} \tag{D.2}$$

An extra step in writing out the first order condition for the dynamic programming problem may be helpful.

$$\begin{aligned}0&=\mathrm{E}_{t}\left(\left(\tilde{y}_{t}+\beta_{t}x_{t}+u_{t}-{}_{t}\bar{u}_{t}\right)\beta_{t}\right)+\rho\,\mathrm{E}_{t}\left(\frac{1}{1-\rho\left(1-\lambda\right)}\cdot\left(1-\lambda\right)\cdot\left(\tilde{y}_{t}+\beta_{t}x_{t}+u_{t}-{}_{t}\bar{u}_{t}+{}_{t+1}\bar{u}_{t+1}\right)\cdot\beta_{t}\right)\\0&=\mathrm{E}_{t}\left(\left(\tilde{y}_{t}+\beta_{t}x_{t}+u_{t}-{}_{t}\bar{u}_{t}\right)\beta_{t}\left(1+\rho\frac{\left(1-\lambda\right)}{1-\rho\left(1-\lambda\right)}\right)+{}_{t+1}\bar{u}_{t+1}\cdot\beta_{t}\left(\rho\frac{\left(1-\lambda\right)}{1-\rho\left(1-\lambda\right)}\right)\right)\end{aligned} \tag{D.3}$$

In deriving the quasi-difference in section 3 it's useful to write out the lag

$$\left(1-\lambda\right)x_{t-1}^{*e}=-\frac{\lambda}{\overline{\beta}}\left[{}_{t-1}\overline{u}_{t-1}\left(1-\lambda\right)+\sum_{n=0}^{t-3}\left(\left[\left(u_{t-2-n}-{}_{t-2-n}\overline{u}_{t-2-n}\right)+{}_{t-2-n}\overline{u}_{t-2-n}\left(1-\lambda\right)\right]\left(1-\lambda\right)^{n}\right)\right] \quad \text{(D.4)}$$