assumes that everyone's income, borrowing constraints, and assets scale proportionately with aggregate income. In contrast, in our model, the extra income is received disproportionately by high income people and the shock leads incomes to rise relative to the households' assets. Both of these features mute the strength of the general equilibrium forces that Werning emphasizes.

Several other recent papers suggest "solutions" to the forward guidance puzzle. Del Negro, Giannoni, and Patterson (2013) argue that the experiment that gives rise to the puzzle is, itself, unreasonable. They argue that it is unreasonable to assume that the central bank really can engender substantial changes in long-term interest rates, which are at the core of why the forward guidance puzzle arises. Carlstrom, Fuerst, and Paustian (2012) and Kiley (2014) show that the magnitude of the forward guidance puzzle is substantially reduced in a sticky information (as opposed to a sticky price) model. This is because the sticky information Phillips curve is less forward looking. Our solution instead yields a Euler equation that is less forward looking than in the standard model. Caballero and Farhi (2014) argue that forward guidance is less effective if the reason why the zero lower bound binds is a shortage of safe assets in the economy, a safety trap, as opposed to a deleveraging or patience shock.

The paper proceeds as follows. Section I explains why forward guidance is so powerful in standard New Keynesian models. Section II presents our incomplete markets model featuring uninsurable idiosyncratic income risk and borrowing constraints. Section III describes our results about the reduced power of forward guidance in our incomplete markets model relative to the standard complete markets models. Section IV concludes.

I. Why Is Forward Guidance So Powerful?

It is useful to start with an explanation for why forward guidance is so powerful in standard monetary models. Consider the basic New Keynesian model as developed, e.g., in Woodford (2003) and Galí (2008). The implications of private sector behavior for output and inflation in this model can be described up to a linear approximation by an intertemporal "IS" equation of the form

$$x_t = E_t x_{t+1} - \sigma\left(i_t - E_t \pi_{t+1} - r_t^n\right), \tag{1}$$

and a Phillips curve of the form

$$\pi_t = \beta E_t \pi_{t+1} + \kappa x_t. \tag{2}$$

Here, x_t denotes the output gap, i.e., the percentage difference between actual output and the natural rate of output that would prevail if prices were fully flexible; π_t denotes inflation; i_t denotes the nominal, short-term, risk-free interest rate; r_t^n denotes the natural real rate of interest, i.e., the real interest rate that would prevail if prices were fully flexible; σ denotes the intertemporal elasticity of substitution; β denotes the subjective discount factor of households; and κ is the slope of the Phillips curve, which is determined by the degree of nominal and real rigidities in the economy. All variables are denoted as percentage deviations from their steady state values.

Suppose for simplicity that the monetary policy of the central bank is given by an exogenous rule for the real interest rate where the real interest rate tracks the natural real rate with some error: $r_t = i_t - E_t\pi_{t+1} = r_t^n + \epsilon_{t,t-j}$. Here $\epsilon_{t,t-j}$ denotes the shock to the short term real rate in period t that becomes known in period $t-j$.[5] Absent any monetary shocks, the real interest rate will perfectly track the natural real rate and both the output gap and inflation will be zero. Suppose we start in such a state, but then the monetary authority announces that the real interest rate will be lower by 1 percent for a single quarter five years in the future, but maintained at the natural real rate of interest in all other quarters (i.e., $\varepsilon_{t+20,t} = -0.01$).[6]

Figure 1 plots the response of output to this shock (assuming for simplicity that $\sigma = 1$). Even though the real interest rate does not change until 20 quarters later, output jumps up by a full 1 percent immediately. Output then stays at this higher level for 20 quarters before falling back to steady state in quarter 21. To understand why output responds in this way, it is important to consider that the shock changes the relative price of consumption between quarters 20 and 21 (since it is the real interest rate in quarter 20 that changes), but leaves the relative price of consumption for any two dates before quarter 20 and any two dates after quarter 20 unchanged. This implies that consumption growth can only deviate from normal in quarter 20. In other words, the response of consumption must be a step function. In addition to this, the level of consumption (and output) is pinned down in the long run by the fact that monetary shocks have no effect on real outcomes in the long run. This implies that consumption (and output) must rise by 1 percent immediately, so that they can fall back to steady state in quarter 21.[7]

The step-function shape of the output response in Figure 1 is determined solely by the Euler equation. The level of consumption, and therefore output, is however determined by the intertemporal budget constraint. In general equilibrium, income rises in response to this type of shock because the level of production increases in response to the increase in demand. This increase in income allows households to consume more initially without reducing consumption after period 20. We can compare this general equilibrium case to the response of a single household holding its own income fixed and also holding the actions of all other agents in the economy fixed (call this the partial equilibrium response). Figure A1 in the Appendix plots the partial equilibrium response. The partial equilibrium response is also a step function, since the same Euler equation applies. The difference is that the increase in consumption over the first 20 quarters will cause the household to run down

[5] Given this specification of monetary policy, the model has a unique solution for which $\lim_{j\to\infty} E_t x_{t+j} = 0$ and inflation is bounded. We could alternatively assume that the monetary authority sets the nominal rate according to the following rule $i_t = r_t^n + \phi\pi_t + \tilde{\epsilon}_{t,t-j}$ and $\phi > 1$. In this case, the model has a unique bounded solution (without the additional restriction that $\lim_{j\to\infty} E_t x_{t+j} = 0$) and there exists a path for $\tilde{\epsilon}_{t,t-j}$ that gives the same solution as the model with monetary policy given by the exogenous path for the real rate we assume. We prefer to describe the monetary policy as an exogenous rule for the real interest rate because this simplifies our exposition substantially.

[6] Conventional specifications of monetary shocks affect real rates in more than a single quarter. But in a linear model the effects of such monetary shocks can be "decomposed" into a simple sum of the effects of changes in real rates at each horizon. In this sense one can think of our 20-quarter experiment as one component of a more complex monetary shock that affects real rates in many quarters.

[7] An alternative way to see this is to solve the intertemporal IS equation—equation (1)—forward to get $x_t = -\sigma\sum_{j=0}^{\infty} E_t\left(i_{t+j} - E_{t+j}\pi_{t+j+1} - r_{t+j}^n\right)$. Notice, that there is no discounting in the sum on the right-hand side of this equation. This implies that the output gap will rise immediately by 1 percent and will stay at that higher level for the next five years and then fall back to zero all at once when the low interest rate period passes.

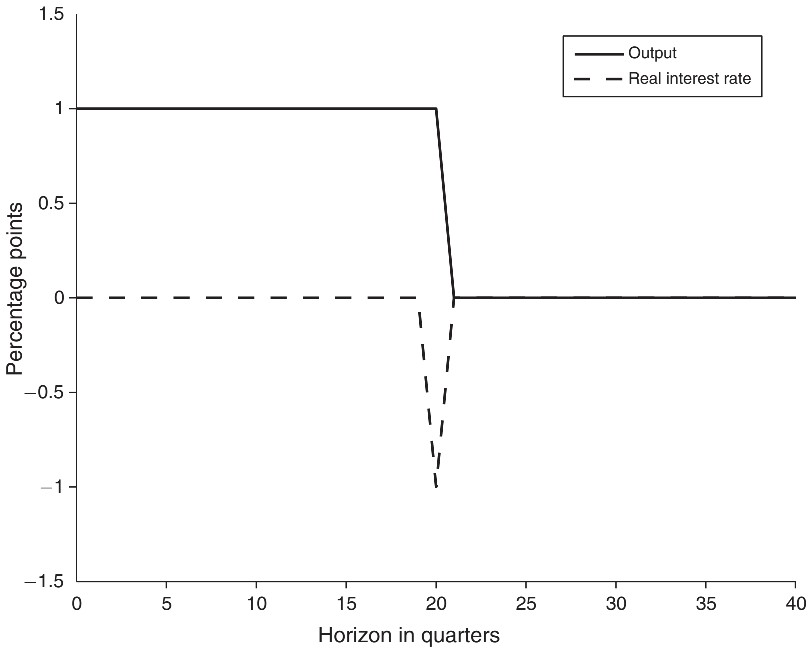

Figure 1. Response of Output to a 1-Quarter Drop in the Real Interest Rate 20 Quarters in the Future

its wealth and imply that consumption going forward after period 20 will be permanently lower (this effect does not occur in general equilibrium due to the offsetting income rise). However, this difference is relatively small, even for a shock 20 quarters out. For this case, the partial equilibrium response of output is 91 basis points rather than the full 100 basis points in general equilibrium.

The logic described above, for forward guidance 20 quarters in the future, applies for forward guidance at any horizon. As a consequence, the further out in the future the forward guidance is, the larger is the *cumulative* response of output. In the New Keynesian model, it is the entire cumulative response of the output gap (albeit with some discounting) that determines the current response of inflation to forward guidance. To see this, it is useful to solve the Phillips curve forward to get

$$\pi_t = \kappa \sum_{j=0}^{\infty} \beta^j E_t x_{t+j}. \tag{3}$$

This equation makes clear that the further in the future is the interest rate that the monetary authority announces it will change, the larger is the current response of inflation. While the response of inflation to a 1 percent change in the current real rate is $\kappa\sigma$, the response of inflation to a 1 percent change in the real rate for one quarter in the infinite future is $\kappa\sigma/(1-\beta)$. If $\beta = 0.99$, the current response of inflation to forward guidance about a single quarter in the infinite future is 100 times larger than the response of inflation to an equally large change in the current real interest rate. Figure 2 plots the response of inflation to forward guidance about interest rates at different horizons relative to the response of inflation to an equally large change in the current real interest rate. We see that the response of inflation to forward

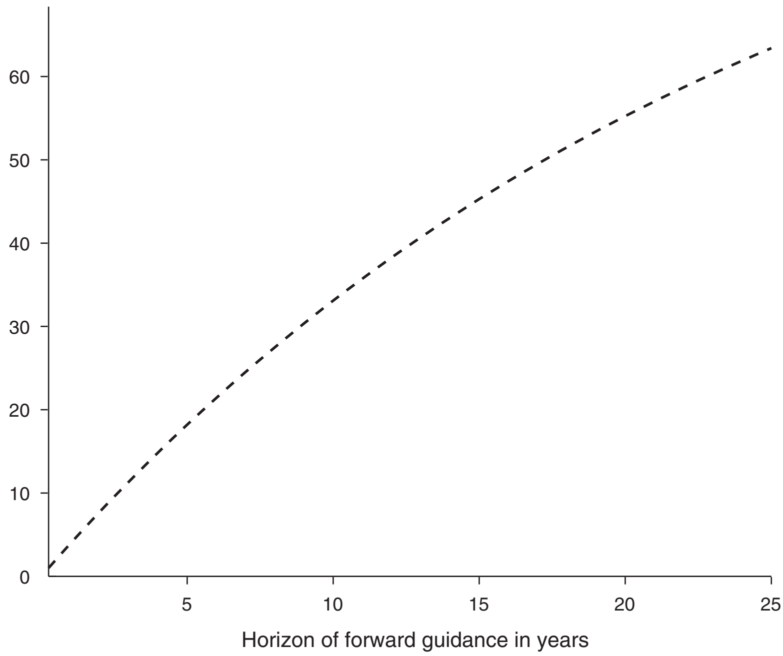

Figure 2. Response of Current Inflation to Forward Guidance about Interest Rates at Different Horizons Relative to Response to Equally Large Change in Current Real Interest Rate

guidance about interest rates five years in the future is roughly 18 times larger than the response of inflation to an equally sized change in the current real interest rate.

To build intuition, we have assumed that there is no endogenous feedback from changes in output and inflation back onto real interest rates. Actual monetary policies are more complicated. In normal times, forward guidance about lower real interest rates in the future may be partly undone by higher real interest rates in the intervening period. On the other hand, when monetary policy is constrained by the zero lower bound on short-term nominal interest rates, the higher inflation associated with forward guidance about future interest rates will actually lower current real interest rates and this will in turn raise current output and inflation further. In this case, the outsized effects of forward guidance we describe above will be further reinforced by subsequent endogenous interest rate movements.

II. An Incomplete Markets Model with Nominal Rigidities

Section I shows that the huge power of far future forward guidance in standard monetary models depends crucially on the prediction of the model that the current response of output to an expected change in real interest rates in the far future (say five years in the future) is equally large as the response of output to a change in the current real interest rate. But is this realistic? With some probability, one will hit a borrowing constraint in the next five years. This effectively shortens one's planning horizon. Also, households that face uninsurable idiosyncratic income risk and borrowing constraints will be wary of running down their wealth to take advantage