Are Nonlinear Methods Necessary at the Zero Lower Bound?*

Alexander W. Richter  Nathaniel A. Throckmorton
Federal Reserve Bank of Dallas  College of William & Mary

August 11, 2016

ABSTRACT

This paper examines the importance of the zero lower bound (ZLB) constraint on the nominal interest rate by estimating three versions of a small-scale New Keynesian model: (1) a nonlinear model with an occasionally binding ZLB constraint; (2) a constrained linear model, which imposes the constraint in the filter but not the solution; and (3) an unconstrained linear model, which never imposes the constraint. The nonlinear model has a higher likelihood in periods when the Fed is constrained and primarily attributes the ZLB to a reduction in demand due to discount factor shocks. In the linear models, the ZLB is due to large contractionary monetary policy shocks, which is at odds with the Fed’s expansionary policy during the Great Recession. Posterior predictive analysis shows the nonlinear model predicts higher output volatility and negative skewness in output and inflation, two features that are essential to match data before and immediately after the Fed lowered its policy rate to zero. In contrast, neither of the linear models predict a change in volatility or skewness when the ZLB binds. We also compare the results from our nonlinear model to the quasi-linear solution based on OccBin. The quasi-linear model fits the data better than the linear models, but it does not generate enough volatility at the ZLB and it predicts that a large policy shock caused the ZLB to bind.

Keywords: Bayesian Estimation; Model Comparison; Zero Lower Bound; Particle Filter

JEL Classifications: C11; E43; E58

*Richter, Research Department, Federal Reserve Bank of Dallas, 2200 N. Pearl Street, Dallas, TX 75201 (alex.richter@dal.frb.org); Throckmorton, Department of Economics, College of William & Mary, P.O. Box 8795, Williamsburg, VA 23187 (nathrockmorton@wm.edu). We thank Ben Keen, Mike Plante, and Todd Walker for suggestions that improved the paper. We also thank Auburn University and especially Bradley Morgan for their support of the Hopper computer cluster. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Dallas or the Federal Reserve System. All remaining errors are our own.
1 Introduction

Long after central banks raise their policy rates, the extended period of near zero rates will remain an important and challenging feature of the data. Of central importance to both current and future work on monetary policy is the question of how to deal with the zero lower bound (ZLB) constraint on the nominal interest rate. Researchers have three options when estimating a model: ignore the constraint, account for the constraint only in the prediction step, or impose the constraint in both the prediction and estimation phases. Recent work uses a variety of models, solution methods, and estimation procedures to study the ZLB period, all of which differ in how they treat the constraint.

This paper examines the importance of the ZLB constraint by estimating three versions of a small-scale New Keynesian model: (1) a nonlinear model with an occasionally binding ZLB constraint; (2) a constrained linear model, which imposes the constraint in the filter but not the solution; and (3) an unconstrained linear model, which never imposes the constraint. The first model serves as our benchmark, because it is the most comprehensive in its treatment of the constraint. In that model, households’ expectations account for the possibility of going to and leaving the ZLB, which depends on the state of the economy as well as current and future shocks. The drawback with this method is that it becomes increasingly costly to evaluate the likelihood function for each draw from the posterior distribution as the size of the model expands. In the second model, households do not account for the constraint in their decision rules, but it is imposed in simulations of the model to prevent negative realizations of the policy rate. By disregarding the effects of the constraint on households’ decisions, it is easy to solve the model using linear methods, which makes it quicker to estimate. The third model completely ignores the constraint when solving and simulating the model. We estimate each model using data on real GDP growth, inflation, and the federal funds rate from 1986Q1 to 2014Q2 by embedding a particle filter into a Metropolis-Hastings algorithm.

A few papers examine the impact of the solution in a calibrated model [Braun et al. (2012); Fernández-Villaverde et al. (2015); Gavin et al. (2015); Nakata (2012)]. Those papers compare the policy functions and impulse responses across a variety of solution methods. While they find meaningful differences between linear and nonlinear models, it is hard to assess their quantitative importance without taking the models to the data. Using artificial data from a small-scale New Keynesian model, Hirose and Inoue (2016) find that ignoring the ZLB during estimation can bias the parameter estimates as the frequency and duration of ZLB events increases. Gust et al. (2016) estimate a medium-scale model with U.S. data to show the empirical implications of the ZLB constraint. These papers stress the importance of using nonlinear estimation techniques by comparing the posterior distributions and impulse responses from the nonlinear and linear models. We build on their work by providing a detailed account of the shocks that kept the economy at the ZLB and showing how well each model fits the data in the ZLB period. To conduct our analysis, we compare the posterior distributions, the marginal data densities, the filtered observables and shocks, the likelihoods, and the posterior predictive distributions of various moments from our three models.

Surprisingly, the posterior distributions and marginal likelihoods from the nonlinear and constrained linear models are similar, but important differences arise in their predictions at the ZLB. The nonlinear model has a higher likelihood in periods when the Fed is constrained and primarily attributes the ZLB to a reduction in demand due to discount factor shocks. In the constrained linear model, the ZLB is due to large contractionary monetary policy shocks, which is at odds with the Fed’s expansionary policy during the Great Recession. The unconstrained linear model has a lower marginal likelihood than the other models, and it predicts even larger policy shocks.
during the ZLB period. A comparison of the posterior predictive distributions shows that the three models match the data equally well in the pre-ZLB period. In the ZLB period, the nonlinear model predicts higher output volatility and negative skewness in output and inflation, two features that are essential to match data before and immediately after the Fed lowered its policy rate to zero. In contrast, neither of the linear models predict a change in volatility or skewness when the ZLB binds. Conditional on 2008Q4, we also find the most common ZLB event is only one quarter in the linear models. In the nonlinear model, it is four quarters, which is consistent with survey data.

Several papers examine the financial crisis and the ensuing Great Recession through the lens of an unconstrained linear model. For example, Ireland (2011) compares the shocks that caused the 1991, 2001, and 2007-2009 recessions using maximum likelihood estimates from an unconstrained, small-scale, linear New Keynesian model. He finds that a major difference between the three recessions is that the most recent recession was plagued by large monetary policy shocks. Suh and Walker (2016) use Bayesian methods to estimate an unconstrained, medium-scale, linear New Keynesian model with financial frictions. They find monetary policy shocks played a major role in explaining the changes in consumption and investment during the Great Recession. Unconstrained linear models are also used by the Fed for policy analysis and forecasting [e.g., Negro et al. (2013)]. Although the ZLB period lasted for several years, our estimates indicate that the Fed was only constrained from 2008Q4 to 2010Q4. If the data used to estimate a model includes that period, it is crucial to use nonlinear estimation techniques. A linear model will lead to incorrect predictions about the causes and consequences of the ZLB because it does not capture the endogenous effects of the ZLB constraint, even if the constraint is imposed when filtering the data.

The nonlinear solution provides the most accurate way to characterize the dynamics just before the ZLB binds and while the central bank is constrained, but it is computationally expensive. An alternative solution method developed by Guerrieri and Iacoviello (2015) is based on a quasi-linear variant of the nonlinear model, where the constraint binds in one regime and is slack in the other regime. The primary advantage of this solution method is that it is as fast as linear methods that ignore the ZLB constraint. The authors have also developed a toolbox, known as OccBin, that is compatible with Dynare and easy to implement. Unfortunately, those benefits come with two major drawbacks. First, households do not account for the expectational effects of going to and leaving the ZLB in their decisions. Second, the economy must return to the regime where the ZLB is slack when simulating the model, which makes it costly to implement a filter that relies on many simulations of the model. Those costs are large enough that it takes longer to draw from the posterior distribution (i.e., solve the quasi-linear model and filter the data) than in our nonlinear model, although that result would reverse if the model contained more state variables and shocks.

We compare the results from our nonlinear model to the quasi-linear solution based on OccBin. The quasi-linear model is better able to capture the increased volatility at the ZLB than the linear model, but it does not generate as much volatility as the nonlinear model, which is necessary to match data during the ZLB period. Furthermore, the quasi-linear model predicts that a large contractionary monetary policy shock caused the ZLB to bind in 2008Q4, just like the linear models. Those differences arise despite the simplicity of our nonlinear model. In models with stronger expectational effects, such as ones with capital or a banking sector, the differences would be larger.

The paper proceeds as follows. Section 2 describes our model. Section 3 outlines our solution method and estimation procedure. Section 4 highlights the differences between the nonlinear and linear models by comparing the posterior distributions and filtered paths. Section 5 compares the dynamics from our nonlinear model to the quasi-linear model using OccBin. Section 6 concludes.
2 STRUCTURAL MODEL

Our model consists of a representative household that has access to a one-period nominal bond, intermediate and final goods producing firms, and a central bank that sets the nominal interest rate. We estimate three versions: (1) the nonlinear model, which imposes the ZLB constraint; (2) a log-linear analogue of the nonlinear model, which removes the constraint from the solution but not the filter; (3) a log-linear analogue of the nonlinear model, which completely removes the constraint.

2.1 HOUSEHOLDS A representative household chooses \{c_t, n_t, b_t\}_{t=0}^{\infty} to maximize expected lifetime utility, \( E_0 \sum_{t=0}^{\infty} \beta_t [\log(c_t - hc_{t-1}^a) - \chi n_t^{1+\eta}/(1 + \eta)] \), where \( \chi > 0, 1/\eta \) is the Frisch elasticity of labor supply, \( c \) is consumption, \( c^a \) is aggregate consumption, \( h \) is the degree of external habit persistence, \( n \) is labor hours, and \( b_t \) is the real value of a privately-issued 1-period nominal bond. \( E_0 \) is an expectation operator conditional on information in period 0, \( \beta_0 \equiv 1, \beta_t = \prod_{j=1}^{t-1} \beta_j \). Just like in other models used to study the ZLB period, \( \beta \) is a time-varying discount factor that follows

\[
\log \beta_t = (1 - \rho_\beta) \log \beta + \rho_\beta \log \beta_{t-1} + \sigma_v v_t, \quad 0 \leq \rho_\beta < 1, \quad v \sim N(0, 1),
\]

where \( \beta \) is the discount factor along the steady state growth path. The choices are constrained by

\[
w_t = \chi n_t^\theta (c_t - hc_{t-1}^a),
\]

\[1 = i_tE_t[q_{t,t+1}/\pi_{t+1}],
\]

where \( q_{t,t+1} \equiv \beta_{t+1}(c_t - hc_{t-1}^a)/(c_{t+1} - hc_t^a) \) is the pricing kernel between periods \( t \) and \( t + 1 \).

2.2 FIRMS The production sector consists of a continuum of monopolistically competitive intermediate goods firms owned by households and a final goods firm. Intermediate firm \( f \in [0, 1] \) produces a differentiated good, \( y_t(f) \), according to \( y_t(f) = z_t n_t(f) \), where \( n(f) \) is the labor hired by firm \( f \) and \( z_t = g_t z_{t-1} \) is technology. The deviations from the steady state growth rate, \( \tilde{g} \), follow

\[
\log g_t = (1 - \rho_g) \log \tilde{g} + \rho_g \log g_{t-1} + \sigma_\varepsilon \varepsilon_t, \quad 0 \leq \rho_g < 1, \quad \varepsilon \sim N(0, 1).
\]

Each intermediate firm chooses its labor to minimize its costs, \( w_t n_t(f) \), subject to its production function. The final goods firm purchases \( y_t(f) \) units from each intermediate firm to produce the final good, \( y_t \equiv \int_0^1 y_t(f)^{(\theta-1)/\theta} df/\theta^{(\theta-1)} \), where \( \theta > 1 \) measures the elasticity of substitution between intermediate goods. It then maximizes dividends to determine its demand function for intermediate good \( f, y_t(f) = (p_t(f)/p_t)^{-\theta} y_t \), where \( p_t = \int_0^1 p_t(f)^{(1-\theta) df}/(1-\theta) \) is the price level.

Following Rotemberg (1982), each intermediate firm pays a cost to adjust its price, \( \text{adj}_t(f) = \varphi[p_t(f)/(\pi p_{t-1}(f))] - 1]^{1/2} y_t/2 \), where \( \varphi > 0 \) scales the cost and \( \pi \) is the gross inflation rate along the steady state growth path. Firm \( f \) chooses its price, \( p_t(f) \), to maximize the expected present value of future dividends, \( E_t \sum_{k=t}^{\infty} q_{t,k} d_k(f) \), where \( q_{t,t} \equiv 1, q_{t,k} \equiv \prod_{j=t+1}^{k} q_{j-1,j} \), and \( d_k(f) = (p_t(f)/p_t) y_t(f) - w_t n_t(f) - \text{adj}_t(f) \). In symmetric equilibrium, the optimality condition implies

\[
\varphi(\hat{\pi}_t - 1) \hat{\pi}_t = 1 - \theta + \theta (w_t/z_t) + \varphi E_t[q_{t,t+1} (\hat{\pi}_{t+1} - 1) \hat{\pi}_{t+1} (y_{t+1}/y_t)],
\]

where \( \hat{\pi}_t \equiv \pi_t/\bar{\pi} \). When \( \varphi = 0, w_t/z_t = (\theta - 1)/\theta \), which is the inverse of the gross price markup.
2.3 Monetary Policy The central bank sets the gross nominal interest rate according to
\[
i_t = \max\{\bar{i}, i_t^\ast\}, \quad i_t^\ast = (i_{t-1}^\ast)^{\rho_1} (\tilde{\pi}_t^{\phi_n} (c_t / (\tilde{g}_t c_{t-1}))^{\phi_c})^{1-\rho_1} \exp(\sigma_\nu \nu_t), \quad 0 \leq \rho_i < 1, \quad \nu \sim N(0,1),
\]
where \(\bar{i}\) is the lower bound, \(i_t^\ast\) is the notional rate, \(\phi_n\) and \(\phi_c\) are the responses to deviations of inflation from target and deviations of consumption growth from its steady state, and \(\tilde{\pi}\) and \(\tilde{\pi}\) are the inflation and interest rate targets, which equal their values along the steady state growth path.

The treatment of the ZLB constraint will influence the estimated monetary policy shocks that explain the data. It is important to note that the monetary policy shock affects the notional rate, not the nominal rate. In the nonlinear model, a positive (negative) policy shock at the ZLB immediately leads to a higher (lower) notional rate and, due to smoothing, higher (lower) than expected future notional rates, which can exceed the ZLB. That shock causes a decrease (increase) in real GDP growth and inflation when the ZLB binds, which affects the model likelihood. In the linear models, monetary policy shocks have similar effects on real GDP growth and inflation regardless of whether the notional rate is negative. If the ZLB is imposed in the filter, then the filter is able to distinguish between the notional and nominal rates. In other words, if the ZLB binds in the data and the model predicts a negative notional rate, then the filter may be able to match the prediction to the data without requiring a positive policy shock to explain the difference, although other properties of the model may still lead to alternative sequences of shocks. The unconstrained linear model, however, cannot distinguish between the two rates. In that case, if the model predicts a negative nominal rate, then positive policy shocks are necessary to make the nominal rate consistent with the data.

2.4 Competitive Equilibrium To make the model stationary, we redefine all of the variables that grow in terms of technology (i.e., \(\tilde{x}_t \equiv x_t / z_t\)). The detrended equilibrium system consists of
\[
\tilde{\lambda}_t = \tilde{c}_t - h \tilde{c}_{t-1} / g_t, \quad \tilde{w}_t = \chi \tilde{y}_t \tilde{\lambda}_t, \quad 1 = i_t E_t[\beta_{t+1}(\tilde{\lambda}_t / \tilde{\lambda}_{t+1}) (1 / (g_{t+1} \tilde{\pi}_{t+1}))], \quad i_t^\ast = (i_{t-1}^\ast)^{\rho_1} (\tilde{\pi}_t^{\phi_n} (g_t \tilde{c}_t / (\tilde{g}_t c_{t-1}))^{\phi_c})^{1-\rho_1} \exp(\sigma_\nu \nu_t), \quad \tilde{c}_t = [1 - \varphi(\tilde{\pi}_t - 1)^2 / 2] \tilde{y}_t \equiv \tilde{y}_t^{d\varphi}, \quad \varphi(\tilde{\pi}_t - 1) \tilde{\pi}_t = (1 - \theta) + \theta \tilde{w}_t + \varphi E_t[\beta_{t+1}(\tilde{\lambda}_t / \tilde{\lambda}_{t+1}) (\tilde{\pi}_{t+1} - 1)] \tilde{\pi}_{t+1} (\tilde{y}_{t+1} / \tilde{y}_t),
\]
the ZLB constraint, and the stochastic processes, which impose the bond market clearing condition, \(b_t = 0\), and the aggregation rule, \(\tilde{c}_t = \tilde{c}_t^\ast\). A competitive equilibrium includes sequences of quantities, \(\{\tilde{\lambda}_t, \tilde{c}_t, \tilde{y}_t\}_{t=0}^\infty\), prices, \(\{w_t, i_t, \tilde{c}_t, \tilde{\pi}_t\}_{t=0}^\infty\), and exogenous variables, \(\{\beta_t, g_t\}_{t=0}^\infty\), that satisfy the detrended system, given the initial conditions, \(\{\tilde{c}_{-1}, i_{-1}^\ast, \beta_0, g_0, \nu_0\}\), and the shocks, \(\{\varepsilon_t, v_t, \nu_t\}_{t=1}^\infty\).

3 Solution Methods and Estimation Procedure

This section concisely describes our solution methods and outlines the estimation procedure applied to all three models. See Plante et al. (2016) for a more detailed description of both algorithms.

3.1 Solution Methods We first solve the log-linear version of our model using Sims (2002) algorithm. Using that solution as an initial conjecture, we then solve the nonlinear model with the policy function iteration algorithm in Richter et al. (2014), which is based on the theoretical work...
on monotone operators in Coleman (1991). This method discretizes the state space and iteratively solves for policy functions that best satisfy the detrended system until a tolerance is met. We use linear interpolation to approximate future variables and follow Rouwenhorst (1995) to integrate.

<table>
<thead>
<tr>
<th>Steady-State Discount Factor</th>
<th>( \hat{\beta} )</th>
<th>0.9984</th>
<th>Nominal Interest Rate Lower Bound</th>
<th>( \hat{\ell} )</th>
<th>1.00017</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frisch Elasticity of Labor Supply</td>
<td>( 1/\eta )</td>
<td>3.0000</td>
<td>Output Growth Measurement Error SD</td>
<td>( \sigma_{\text{me}, g} )</td>
<td>0.00194</td>
</tr>
<tr>
<td>Elasticity of Substitution between Goods</td>
<td>( \theta )</td>
<td>6.0000</td>
<td>Inflation Rate Measurement Error SD</td>
<td>( \sigma_{\text{me}, \pi} )</td>
<td>0.00075</td>
</tr>
<tr>
<td>Steady-State Labor</td>
<td>( \hat{\bar{n}} )</td>
<td>0.33</td>
<td>Nominal Interest Rate Measurement Error SD</td>
<td>( \sigma_{\text{me}, i} )</td>
<td>0.00206</td>
</tr>
</tbody>
</table>

Table 1: Calibrated parameters and measurement error standard deviations.

3.2 Estimation Procedure We estimate our models using quarterly data on per capita real GDP \((Y/N)\), the GDP price deflator \((P)\), and the federal funds rate \((i)\) from 1986Q1 to 2014Q2. The matrix of observables is \( x_{\text{data}} \equiv [\log(Y_t/N_t) - \log(Y_{t-1}/N_{t-1}), \log(P_t/P_{t-1}), \log(i_t)] \). We calibrate five of the parameters (table 1). The steady-state discount factor, \( \hat{\beta} \), is set to 0.9984, which equals \((1/T) \sum_{k=1}^{T} g_k \pi_k / i_k\) where \( T \) is the sample size and \( g_k \) is Fernald’s (2012) utilization-adjusted growth rate of technology. The preference parameter, \( \chi \), is set so steady-state labor equals 1/3 of total hours. The elasticity of substitution, \( \theta \), equals 6, which implies a 20% average markup. The lower bound on the policy rate, \( \hat{\ell} \), is set to 1.00017, which is the minimum federal funds rate in the ZLB period. The Frisch elasticity of labor supply, \( 1/\eta \), is set to 3 following Peterman (2016).

We use a random walk Metropolis-Hastings algorithm with a particle filter to evaluate the likelihood of the posterior distribution following Fernández-Villaverde and Rubio-Ramírez (2007). Similar to Herbst and Schorfheide (2016), we adapt the filter to include information from the current period, which helps the model match outliers that occurred during the Great Recession. The filter uses 40,000 particles and systematic resampling with replacement following Kitagawa (1996). We convert the predictions of the linear models to levels, so we can apply the same filter to all three models. Given the simulated paths from each model, we transform the predictions for real GDP, inflation, and the policy rate according to \( x_{t, \text{data}} = x_{t, \text{model}} + \xi_t \), where \( \xi_t \sim N(0, \Sigma) \) is a vector of measurement errors and \( \Sigma = \text{diag}([\sigma_{\text{me}, g}^2, \sigma_{\text{me}, \pi}^2, \sigma_{\text{me}, i}^2]) \). The standard deviation (SD) of each measurement error is set to 10% of the variance of the data (table 1). We obtain 100,000 draws from the posterior distribution and keep every 100th draw, so our analysis is based on 1,000 draws.

The entire algorithm is programmed in Fortran using Open MPI and executed on a cluster. We parallelize the nonlinear solution by distributing the nodes in the state space across the available processors. To increase the accuracy of the filter, we calculate the posterior likelihood on each available processor and evaluate whether to accept or reject a given draw based on the median likelihood. Therefore, the effective number of particles equals 40,000 times the number of processors.

4 Model Comparison

This section evaluates the performance of the nonlinear, constrained linear, and unconstrained linear models across a number of dimensions. Specifically, we show the posterior distributions, the predicted observables and shocks, the impulse responses, a time series of the filter densities, the posterior predictive distributions of three key moments, and the predicted duration of ZLB events.

4.1 Posterior Distributions The first three columns of table 2 show the parameters and prior distributions, which are relatively diffuse and broadly consistent with other papers that use
Bayesian methods. The remaining columns compare the posterior means and 90% credible sets for our three models. Interestingly, most of the posterior means for the nonlinear model are well within the credible sets of both linear models, and there is little difference in their credible sets. One exception is the persistence of the discount factor. The tail of the credible set in the nonlinear model barely includes the posterior mean from the linear models, which is important since it affects the frequency and duration of ZLB events. There are also differences in the two policy parameters.

Given the similarities in the posterior distributions, it might be tempting to estimate the constrained linear model, either on the entire sample or a subsample of the data that does not include the ZLB period, and then solve and simulate a nonlinear version of the model conditional on the mean parameter estimates [e.g., Christiano et al. (2015); Cuba-Borda (2014)]. While that approach is much less computationally expensive and will likely provide a decent approximation in a small-scale New Keynesian model, larger differences between the posterior means will arise in more complicated models. For example, Gust et al. (2016) compare the posterior distributions from a constrained linear and nonlinear version of a medium-scale model with capital. Similar to our results, they find that many of the parameter estimates are similar across the two models, but there are large differences in the parameters governing process for the marginal efficiency of investment as well as important differences in the SDs of the other shocks and the monetary policy parameters.

We also compute the marginal log-likelihood of each model using Geweke’s (1999) harmonic

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Dist</th>
<th>Mean (SD)</th>
<th>Nonlinear</th>
<th>Constrained Linear</th>
<th>Unconstrained Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \varphi )</td>
<td>Gam</td>
<td>80.00 (20.00)</td>
<td>96.80137 (67.71867, 131.85091)</td>
<td>92.32496 (60.60734, 127.29165)</td>
<td>91.19524 (60.88623, 129.21631)</td>
</tr>
<tr>
<td>( h )</td>
<td>Beta</td>
<td>0.500 (0.200)</td>
<td>0.44428 (0.30733, 0.57745)</td>
<td>0.44718 (0.29442, 0.58546)</td>
<td>0.44579 (0.30097, 0.58384)</td>
</tr>
<tr>
<td>( \rho_g )</td>
<td>Beta</td>
<td>0.500 (0.200)</td>
<td>0.20064 (0.06547, 0.37805)</td>
<td>0.19063 (0.06085, 0.36433)</td>
<td>0.18582 (0.05543, 0.36851)</td>
</tr>
<tr>
<td>( \rho_\beta )</td>
<td>Beta</td>
<td>0.500 (0.200)</td>
<td>0.90245 (0.87001, 0.92958)</td>
<td>0.92920 (0.88436, 0.96588)</td>
<td>0.92326 (0.88096, 0.96010)</td>
</tr>
<tr>
<td>( \rho_i )</td>
<td>Beta</td>
<td>0.500 (0.200)</td>
<td>0.81158 (0.75375, 0.86060)</td>
<td>0.83541 (0.78091, 0.87712)</td>
<td>0.83846 (0.78707, 0.88103)</td>
</tr>
<tr>
<td>( \sigma_\varepsilon )</td>
<td>IGam</td>
<td>0.010 (0.010)</td>
<td>0.00968 (0.00738, 0.01241)</td>
<td>0.00981 (0.00743, 0.01274)</td>
<td>0.00975 (0.00752, 0.01239)</td>
</tr>
<tr>
<td>( \sigma_\nu )</td>
<td>IGam</td>
<td>0.010 (0.010)</td>
<td>0.00215 (0.00159, 0.00286)</td>
<td>0.00216 (0.00161, 0.00290)</td>
<td>0.00197 (0.00147, 0.00261)</td>
</tr>
<tr>
<td>( \sigma_\nu )</td>
<td>IGam</td>
<td>0.010 (0.010)</td>
<td>0.00199 (0.00148, 0.00261)</td>
<td>0.00187 (0.00139, 0.00242)</td>
<td>0.00181 (0.00137, 0.00232)</td>
</tr>
<tr>
<td>( \phi_{\nu} )</td>
<td>Normal</td>
<td>2.500 (1.000)</td>
<td>4.06383 (3.3317, 4.90267)</td>
<td>4.12432 (3.30566, 4.96845)</td>
<td>3.81212 (3.06638, 4.66968)</td>
</tr>
<tr>
<td>( \phi_g )</td>
<td>Normal</td>
<td>1.000 (0.400)</td>
<td>1.49057 (1.12702, 1.87727)</td>
<td>1.37259 (1.00482, 1.77689)</td>
<td>1.25659 (0.88962, 1.67115)</td>
</tr>
<tr>
<td>( \ddot{g} )</td>
<td>Normal</td>
<td>1.004 (0.001)</td>
<td>1.00376 (1.0026, 1.00489)</td>
<td>1.0037 (1.00252, 1.00493)</td>
<td>1.00368 (1.00251, 1.00478)</td>
</tr>
<tr>
<td>( \ddot{\pi} )</td>
<td>Normal</td>
<td>1.006 (0.001)</td>
<td>1.00622 (1.00556, 1.00683)</td>
<td>1.00621 (1.00557, 1.00687)</td>
<td>1.00608 (1.00545, 1.00673)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Dist</th>
<th>Mean (SD)</th>
<th>Nonlinear</th>
<th>Constrained Linear</th>
<th>Unconstrained Linear</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \log(ML) )</td>
<td></td>
<td></td>
<td>1506.66</td>
<td>1506.45</td>
<td>1499.75</td>
</tr>
</tbody>
</table>

Table 2: Prior distributions, means, standard deviations, and credible sets of the estimated parameters.
mean estimator. The exponential of the difference between any two of those values is the Bayes factor, which is equivalent to the posterior odds ratio since our three models have the same prior distributions. The Bayes factor between the nonlinear and constrained linear models indicates that neither model provides a superior empirical fit over the entire sample, although, as we will show, there are meaningful differences in likelihoods during the ZLB period. These models, however, are superior to the unconstrained linear model, which has a smaller likelihood over the entire sample.

![Graph showing Real GDP Growth Rate, Inflation Rate, and Nominal Interest Rate](image)

Figure 1: Median filtered observables minus the data in annual percentage points. The vertical dashed line is 2008Q4.

4.2 Filtered Observables and Shocks Despite similar posterior distributions, the particle filter shows the nonlinear and linear models fit the data differently and lead to competing explanations of how the economy arrived and stayed at the ZLB. Figure 1 shows the level differences between the median filtered observables and the data. The difference is shown in annual percentage points, so 0.1 means the filtered observable is 0.1 percentage points higher than the data. The vertical line in 2008Q4 indicates when the federal funds rate first fell below 0.25%. In that quarter, the nonlinear model provides a closer fit to the data on real GDP growth and inflation than the linear models. Also, both linear models over-predict the policy rate, while the nonlinear model under-predicts the rate. The nonlinear model continues to better match output and inflation until 2010Q4, although the differences become smaller over time. After 2010Q4, the paths implied by the nonlinear and constrained linear models are very similar. The path of the nominal interest rate from the unconstrained linear model, however, is persistently higher than in the other models.
The models differ most in their predictions of the shocks during the ZLB period. Figure 2 plots the median filtered shocks from the linear models minus the shocks from the nonlinear model. The differences are shown relative to the posterior mean estimate of the shock SD in the nonlinear model, so 2 means a linear model predicted a shock 2SDs larger than in the nonlinear model. In 2008Q4, the differences are stark. Though the nonlinear and constrained linear models predict similar discount factor shocks, the constrained linear model predicts that a larger negative technology shock (−0.2SD) and a much larger monetary policy shock (+1.4SD) are necessary to explain the data in 2008Q4. The reason for staying at the ZLB is also quite different. In 2009Q2, the constrained linear model predicts a larger discount factor (+0.4SD) and monetary policy shock (+0.8SD) than the nonlinear model. The unconstrained linear model also predicts implausibly large policy shocks from 2008Q4 to 2009Q4, but the discount factor shocks are slightly different.

The differences in the filtered paths stem from the negative notional rate the nonlinear model predicts from 2008Q4 to 2010Q4 (figure 3). During that period, the Fed could not lower its policy rate even though it would have preferred to set the rate to −5.2% (−1.3% quarterly) in 2009Q3 because of the large contraction in real GDP growth and the negative inflation rate. In the linear models, households believe the central bank will set a negative policy rate in a severe recession regardless of whether the constraint is imposed in the particle filter, which means the shocks produce different dynamics than in the nonlinear model. After 2010Q4, the nonlinear model predicts a near-zero notional rate, so the filtered paths are similar across the two models. Gust et al. (2016)
estimate a similar path for the notional rate, despite the differences between our nonlinear models. The key takeaway from the path of the notional rate is that the Fed was only constrained for about two years, even though the federal funds rate was still below 0.25% at the end of our sample. Had the Fed been constrained for a longer period, the differences in the shocks would have been larger.

4.3 Generalized Impulse Responses Generalized impulse responses (GIRFs) help us understand why our models produce different results when the ZLB binds. To compute a GIRF, we follow the procedure in Koop et al. (1996). We first calculate the mean of 10,000 simulations of a given model, conditional on random shocks in every quarter. We then calculate a second mean from another set of 10,000 simulations, but this time the shock in the first quarter is replaced with the shock of interest. The GIRF reports the difference between the two mean paths in a given model. Figure 4 plots the responses to a 2SD positive monetary policy (top row), technology (middle row), and discount factor (bottom row) shock in our three models. To compare the responses, we initialize each simulation at the filtered state corresponding to 2009Q1 in the nonlinear model.

A positive technology shock decreases the marginal cost of production, which generates a tradeoff between output and inflation just like a typical supply shock. The competing effects on real GDP growth and inflation cause technology shocks to have a relatively small impact on the notional rate and whether the ZLB binds compared to discount factor shocks. Therefore, the responses are similar across our models. A positive monetary policy shock, however, directly affects the notional rate. In the linear models, the nominal and notional rates are equal so the shock affects households’ decisions even when the ZLB binds. In the nonlinear model the shock only affects the economy if it is large enough to push the nominal interest rate above its ZLB. Since the economy begins in a deep recession and several simulations never exit the ZLB, the monetary policy shock has a much smaller effect on real GDP growth and inflation in the nonlinear model than in the linear models.

A large discount factor indicates that households have a strong desire to save. Elevated savings depresses demand, which reduces output, inflation, and the notional interest rate. In the nonlinear model, any further reduction in expected inflation is offset by an equal increase in the real interest rate since the nominal interest rate is constrained. The higher real rate raises the cost of current consumption, which further lowers demand relative to the linear model such that the responses of real GDP growth and inflation are more than twice as large. Both linear models must compensate for the damped responses to discount factor shocks at the ZLB, which explains why the linear models need a larger negative technology shock and a larger positive policy shock to explain the data in 2008Q4 and both larger discount factor and larger policy shocks to explain the data in 2009.
4.4 Filter Densities  The predicted notional rate and the volatility of real GDP growth and inflation that occurred at the ZLB affects each model’s ability to fit the data. One measure of the empirical fit is shown in figure 5, which plots a time series of the filter density in both linear models minus the density in the nonlinear model. The linear model’s fit is poorest when the notional rate is negative and declining since it cannot generate the heightened volatility necessary to explain the severity of the crisis. For example, the differences in 2008Q4 imply a Bayes factor of about 8, which is a large value even over the entire sample. Eventually, both linear models fit the data.
marginally better than the nonlinear model once the notional rate begins to rise in 2009Q3, which is likely due to the negative notional rate creating persistent volatility in the nonlinear model well after the effects of the crisis subsided. The higher likelihood coming from the linear models, however, is not significant enough to generate a noticeable difference in how well the observables fit the data. Overall, the filtered time series demonstrate that it is crucial to use nonlinear estimation techniques if the sample and period of interest includes quarters when the notional rate is well below zero.

4.5 Posterior Predictive Analysis Posterior predictive analysis provides another way to identify the strengths and weaknesses of each model. We compare moments of interest in the pre-ZLB period (1986Q1-2008Q3) and the ZLB period (2008Q4-2014Q2) to their posterior predictive distributions in each model following the methods in Geweke (2005) and Faust and Gupta (2012). For a given model, we conduct 10,000 stimulations for each draw from the posterior distribution. We initialize the simulations in the pre-ZLB period with a state vector drawn from the model’s ergodic distribution. Each simulation has the same length as the data, and we condition on periods when the ZLB does not bind. To compute the distributions in the ZLB sample, we initialize each simulation at the filtered state corresponding to 2008Q4. We then condition on simulations with a ZLB event that is at least 16 quarters, so the sample moments are based on ZLB events that have a similar duration as the data. It also guarantees that we have a large enough sample size to compute moments in the ZLB period. Given the simulated paths, we calculate time averages of the statistics of interest in each sample and then compute the means and quantiles across the simulations, so the distributions account for the uncertainty surrounding both the shocks and the parameter estimates.

Table 3 shows the mean and 90% credible sets for the mean, SD, and skewness of real GDP growth and inflation in the pre-ZLB period (top panel) and ZLB period (bottom panel) of each model. Not surprisingly, the three models produce similar distributions in the pre-ZLB period. The distributions for the means and SDs of real GDP growth and inflation are wide but their means are near their sample averages. The skewness in the pre-ZLB period, however, is very unlikely to occur in these models. In the data, there is downside risk to real GDP growth and upside risk to inflation, but there is virtually no asymmetry in any of the models when the ZLB does not bind.

In ZLB period, mean real GDP growth and inflation are lower, the SD of real GDP growth is higher, the SD of inflation is slightly lower, and both real GDP growth and inflation are negatively skewed in the data. The nonlinear model better matches these features than the linear models in several ways. First, average real GDP growth is about one-third of its value in the pre-ZLB period, whereas it is only slightly lower in the linear models. Second, real GDP growth is far more volatile than in the pre-ZLB period. In contrast, neither of the linear models predict any change in volatility. It appears that the nonlinear model overpredicts the volatility of real GDP growth, but when we restrict the sample to periods when the notional rate was below zero, the volatility increases to 4.18, which is close to the model. Third, in the nonlinear both real GDP growth and inflation are negatively skewed, while the linear model generates almost no skewness. The nonlinear model does not perform as well as the linear model along two dimensions: it overpredicts the decline in the average inflation rate and it predicts a modest increase rather than a slight decrease in the volatility of inflation. The linear models predict modest declines in the mean and SD of inflation. While the linear model is not capable of generating higher volatility or skewness at the ZLB, it is possible to add features to the nonlinear model that would better match the mean and SD of inflation at the ZLB. Therefore, we believe these results provide support for the nonlinear model.

Figure 6 plots the posterior predictive distribution of ZLB event durations. The unconditional
durations (top panel) are similar across the three models. The most frequent ZLB event is only one quarter and the average duration ranges from 3.1 quarters in the nonlinear model to 4.1 quarters in the constrained linear model. The higher mean and slightly higher likelihood of a lengthy ZLB event in the linear model is mostly due to the higher estimated persistence of the discount factor. The durations conditional on the filtered state in 2008Q4 (bottom panel), however, are significantly different. In both linear models, the most frequent ZLB event remains one quarter, whereas the most common ZLB event duration in the nonlinear model is between 3 and 4 quarters. Also, the average duration in the nonlinear model increases to 5.5 quarters and lengthy ZLB events become more common than in the linear models, despite the differences in the discount factor persistence.

The durations implied by the nonlinear model are far more consistent with survey data. For example, in the January 2009 Blue Chip survey of financial forecasters, 59% of the forecasters predicted the federal funds rate would exceed 0.25% after 3 or 4 quarters (26% predicted 3 quarters and 33% predicted 4), but only 9% predicted that it would exceed 0.25% in 2 or fewer quarters. In the nonlinear model, 37.7% of ZLB events are six or more quarters, whereas 26% of survey respondents thought the federal funds rate would remain below 0.25% for six or more quarters. Evidently, most forecasters expected the Fed to maintain its zero interest rate policy for longer than a couple quarters but not an extended period of time, despite the severity of the Great Recession.

Another important determinant of the dynamics in each model is the frequency of ZLB events. A lower (higher) frequency means households place a smaller (larger) weight on the ZLB in expectation, so the more volatile dynamics at the ZLB have a smaller (larger) impact on households’ decisions away from the ZLB. Unfortunately, there is no reliable measure of this statistic in the

<table>
<thead>
<tr>
<th>Real GDP Growth ($\bar{y}_t^{gdp}$)</th>
<th>Inflation Rate ($\pi_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td>Data</td>
<td>0.21</td>
</tr>
<tr>
<td>Nonlinear</td>
<td>0.57</td>
</tr>
<tr>
<td>(−1.20, 2.32)</td>
<td>(3.12, 6.17)</td>
</tr>
<tr>
<td>Constrained Linear</td>
<td>1.16</td>
</tr>
<tr>
<td>(−0.58, 2.88)</td>
<td>(1.76, 3.56)</td>
</tr>
<tr>
<td>Unconstrained Linear</td>
<td>1.24</td>
</tr>
<tr>
<td>(−0.46, 2.98)</td>
<td>(1.76, 3.43)</td>
</tr>
</tbody>
</table>

(a) Pre-ZLB period (1986Q1-2008Q3).

<table>
<thead>
<tr>
<th>Real GDP Growth ($\bar{y}_t^{gdp}$)</th>
<th>Inflation Rate ($\pi_t$)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean</strong></td>
<td><strong>SD</strong></td>
</tr>
<tr>
<td>Data</td>
<td>1.57</td>
</tr>
<tr>
<td>Nonlinear</td>
<td>(0.68, 2.47)</td>
</tr>
<tr>
<td>Constrained Linear</td>
<td>1.50</td>
</tr>
<tr>
<td>(0.59, 2.41)</td>
<td>(2.11, 3.12)</td>
</tr>
<tr>
<td>Unconstrained Linear</td>
<td>1.49</td>
</tr>
<tr>
<td>(0.60, 2.38)</td>
<td>(2.12, 3.12)</td>
</tr>
</tbody>
</table>

(b) ZLB period (2008Q4-2014Q2).

Table 3: Comparison between various moments of interest in the data and their posterior predictive distributions in each model. The values shown in parentheses are 90% credible sets, and all of the values are annualized net rates.
data. The data contains one ZLB event that lasted for 23 or 20% of the 114 quarters in our sample. With a longer sample, the frequency would decline in the data, but the model-implied distributions would likely remain unchanged. Survey data is also unreliable because of how quickly the Fed lowered its policy rate. Therefore, there is no clear way to determine the frequency of ZLB events.

5 NONLINEAR VS. QUASI-LINEAR MODEL

We focus on how the nonlinear model compares with the linear models because it provides the most accurate predictions of the dynamics at the ZLB. A popular alternative is to solve a quasi-linear version of the nonlinear model using the OccBin toolbox developed by Guerrieri and Iacoviello (2015). OccBin is accompanied by a number of examples including a small-scale New Keynesian model with a ZLB constraint. For that model, the authors find some dynamic differences between the nonlinear solution and their quasi-linear solution. The question remains whether those differences are empirically relevant. We first demonstrate those differences with our New Keynesian model and then examine the implications for the shocks that explain the data in the ZLB period.

Figure 7 shows the impulse responses of the observables to each shock in the nonlinear model (solid line) and the quasi-linear model (dashed line). Those responses are initialized at the median filtered state corresponding to 2009Q1 from the nonlinear model. As one might expect, the quasi-linear model does a better job matching the dynamics of the nonlinear model than the linear model, but there is still a key difference. Except for the monetary policy shock, the quasi-linear model does
not produce as much volatility when the ZLB binds as the nonlinear model, which is most apparent in the responses of inflation and the notional rate to a technology and discount factor shock. That means the model must explain the data with larger and different shocks than the nonlinear model.

Figure 8 shows the median filtered shocks in the quasi-linear model minus the shocks in the nonlinear model. The data is filtered conditional on the posterior mean parameterization from the nonlinear model and the shocks are normalized by their mean standard deviations. In 2008Q4, the quasi-linear model, like the linear models, requires a large discount factor (+1.2SD) and monetary policy (+0.8SD) shock to explain the data. That result is consistent with the fact that the quasi-linear solution does not capture the expectational effects of going to and remaining at the ZLB, which lead to a reduction in real GDP growth and inflation in the nonlinear model. Throughout 2009, the quasi-linear model predicts larger positive discount factor shocks just like the linear models, but similar-sized monetary policy shocks when compared to the nonlinear model. The reason is shown in figure 7. The quasi-linear model does a good job matching the nonlinear responses to a policy shock but a poorer job matching the responses to a discount factor shock. Also, the discount factor is the main shock that causes the ZLB to bind, which influences the expectation of going to and leaving the ZLB. Since the quasi-linear model does not capture those expectational effects, the filtered discount factor shocks are the source of the largest differences from the nonlinear model.

Our results show that the quasi-linear model does not generate enough volatility at the ZLB.
and it counterfactually predicts that a large monetary policy shock caused the ZLB to bind in 2008Q4. The primary advantage of using the quasi-linear model over the nonlinear model is that it is much quicker to solve. For example, with 64 processors it typically takes about 3 seconds to solve the nonlinear model with our the policy function iteration algorithm, whereas it takes less than one second to solve the quasi-linear model. To make matters worse, the nonlinear solution time will increase exponentially with the size of the model (i.e., the number of state variables and shocks), whereas the quasi-linear solution time is nearly independent of the size of the model. Unfortunately, the quasi-linear model is costly to simulate. Given the nonlinear solution, it takes roughly 6 seconds to simulate the model with a particle filter. In contrast, filtering the quasi-linear model with 64 processors and the same number of particles as the nonlinear model takes about 30 seconds. Furthermore, parallelizing the simulations prevents us from distributing multiple runs of the particle filter across the available processors. Therefore, the effective number of particles is 1/64th of the nonlinear estimation. It is possible to integrate other filters with OccBin, but those methods will come with their own drawbacks in terms of speed and accuracy. It may be appealing to use OccBin for large models, but researchers should be cautious of its predictions at the ZLB.

6 Conclusion

This paper analyzes the importance of incorporating the ZLB constraint into households expectations. To conduct our analysis, we compare the posterior distributions based on three models that differ in their treatment of the constraint: (1) a nonlinear model, which includes an occasionally binding ZLB constraint; (2) a constrained linear model, which imposes the constraint in the filter but not the solution; and (3) an unconstrained linear model, which never imposes the constraint.
We find that our models generate similar posterior distributions when estimated with U.S. data that includes the ZLB period. However, the models provide qualitatively different explanations for why the U.S. economy arrived and stayed at the ZLB. The nonlinear model primarily attributes the ZLB to a reduction in household demand due to discount factor shocks, while the linear model incorrectly predicts that the ZLB stems from large contractionary monetary policy shocks. Posterior predictive analysis shows the nonlinear model is better able to match the increased volatility and skewness that occurred during the ZLB period. Interestingly, those differences occur even though we use a small-scale model that only includes three shocks and two endogenous state variables.

We also compare the predictions from our nonlinear model to its quasi-linear analogue using OccBin. The dynamics in the quasi-linear model are closer to those in the nonlinear model than the linear model, but it still does not generate the volatility that is necessary to match the data when the ZLB binds. As a consequence, the quasi-linear model predicts much larger discount factor and monetary policy shocks at the ZLB. These results demonstrate that it is important to use nonlinear solution and estimation techniques when analyzing the ZLB, despite their computational burden.

REFERENCES


