NBG Working Papers

WP 01/2019

Solving non-linear dynamic models (more) efficiently: *application to a simple monetary policy model*

by Shalva Mkhatrishvili, Douglas Laxton, Davit Tutberidze, Tamta Sopromadze, Saba Metreveli, Lasha Arevadze, Tamar Mdivnishvili and Giorgi Tsutskiridze

საქართველოს ეროვნული ბანკი
National Bank of Georgia

Solving non-linear dynamic models (more) efficiently: *application to a simple monetary policy model*[*]

Shalva Mkhatrishvili,[†‡] Douglas Laxton,[⊗] Davit Tutberidze,[‡] Tamta Sopromadze,[‡] Saba Metreveli,[‡] Lasha Arevadze,[‡] Tamar Mdivnishvili[‡] and Giorgi Tsutskiridze[‡]

October 2019

Abstract

There has been an increased acceptance of non-linear linkages being the major driver of the most pronounced phases of business and financial cycles. However, modelling these non-linear phenomena has been a challenge, since existing solutions methods are either efficient but not able to accurately capture non-linear dynamics (e.g. linear methods), or accurate but quite resource-intensive (e.g. stacked system or stochastic Extended Path). This paper proposes two new solution approaches that try to be accurate enough and less costly. Moreover, one of those methods lets us do Kalman filtering on non-linear models in a non-linear way, which is also important for this kind of models, in general, to be more policy-relevant. Impulse responses, simulations and Kalman filtering exercises show the advantages of those new approaches when applied to a simple, but strongly non-linear, monetary policy model.

JEL Codes: C60, C61, C63, E17

Keywords: Non-linear dynamic models; Solution methods; Monetary policy

[*] We would like to thank Gerhard Rünstler (European Central Bank) and all the participants of the 22nd Central Bank Macroeconomic Modeling Workshop 2019 for useful comments and discussion. We would also like to thank Zviad Zedginidze (International Monetary Fund), Archil Imnaishvili (National Bank of Georgia) and Macroeconomics and Statistics, and Financial Stability Departments at the NBG for their feedback. All errors and omissions remain with the authors.

[†] *Corresponding author*: Shalva Mkhatrishvili, National Bank of Georgia, Head of Macroeconomic Research Division (e-mail: shalva.mkhatrishvili@nbg.gov.ge)

[‡] Macroeconomic Research Division; National Bank of Georgia

[⊗] Adjunct Professor of Economics at NOVA School of Business and Economics; Director of Saddle Point Research and The Better Policy Project

Contents

1 Introduction

The importance of solving and estimating non-linear dynamic models has increased significantly after the Global Financial Crisis (2007-2009). One of the important reasons has been heightened attention to macro-financial linkages, which are non-linear in nature (Claessens and Kose, 2018), e.g. accumulation of excessive leverage, large exchange rate fluctuations, house price booms and busts, state (e.g. balance sheet size) dependence of economic agents' decisions. Macro-financial linkages have also been important factor in emerging market economies as well. For example, financial dollarization may change models' propagation mechanism depending on the size of indebtedness in foreign currency. Currency mismatches may also give rise to self-fulfilling banking and macroeconomic crises (Ize and Levy-Yeyati, 2006). Understanding all of these issues requires non-linear models, while non-linear models require practically usable non-linear solution and estimation methods. This is exactly the motivation of this paper.

Solving/simulating non-linear dynamic stochastic models is usually done using either approximate but efficient approaches (e.g. first order approximation around steady states) or much more resource-intensive but arbitrarily accurate methods (e.g. stacked systems or stochastic Extended Path). Improving upon efficiency/accuracy trade-off has been an important area of research on solution methods (Fernndez-Villaverde et al, 2016). This paper tries to propose and investigate two new approaches that could make the solution more efficient *and/or* accurate.

The first method is about making the stacked system or stochastic-version Extended Path (EP) approach more efficient with about the same level of accuracy. This is done by acknowledging that big part of what makes the EP more resource-intensive is the presence of forward-looking (or jump) variables. Namely, for given values of time-$(t-1)$ variables, in order to calculate time-t values we need time-$(t+1)$ values, but in order to calculate the latter we, in turn, need time-$(t+2)$ values and so on. This way we end up with a very big system of non-linear equations (two-point boundary value problem). However, if we could substitute these forward variables (or expected values) with something that both retains many features of forward-lookingness but simplifies the solution problem, then the approach would be more efficient (meaning that the nonlinear system would be of smaller size). This can be done by combining linear solution methods with the EP *only* when it comes to jump variables. In particular, we can always linearize the model and express those time-$(t+1)$ variables in terms of time-t variables (and time-$(t+1)$ shocks, if they are anticipated), instead of relying on time-$(t+2)$ and further ahead values. With this substitution the solution becomes much easier. Since the approach relies on combining more traditional linear solution methods with the non-linear methods, one could call it a Combined Path (CP). One can also think of it as calculating the single dots of a path individually but consistently (recursively), as opposed to the EP, and then combining them to get the whole path of the system.

One of the drawbacks of the CP is that it introduces one source of approximation error. Namely, if expected values in the system, and not the current or predetermined variables, bear most of the non-linear features, then approximating those expectations in a linear fashion may yield inaccurate results. But one may argue that non-linearities often come from predetermined variables (like, balance sheets of banks, corporates or households, credibility of central banks, capital, net worth, etc.). Furthermore, when standard EP has another source of approximation error (the presence of terminal point), the CP approach gets rid of it. Simulations below show that, indeed, introducing new source of approximation error is largely compensated by getting rid of another one. At the same time those simulations also show that this almost-the-same level of accuracy comes at a much lower cost. In particular, the time needed to run the simulations using the CP approach is about six times less than the time needed to do the same simulations using the EP. This is exactly what we are after - improving efficiency while preserving the accuracy.

The second method is about using linear solution methods in a way that results in accuracy comparable to that of non-linear solution methods. This approach has one additional, and probably more important, advantage as it lets us use Kalman filtering techniques easily on a non-linear model. Indeed, likelihood estimation for non-linear models, which is very important for models to be useful in practice, has been difficult with other solution methods (An and Schorfheide, 2007). And solving this problem, even partly, is a big step forward. In addition, this approach does not involve any optimization algorithm (like Newton's method) which is the major reason for solution methods being resource-intensive. The way this method works is based on a simple, but very important, observation: when we talk about linearization around the steady state of the model (as is usually done), one should remember that here the crucial word is "linearization" and not "around the steady state". If the economy is close to the steady state then that kind of linearization may be appropriate. But if the economy is far away from the steady state, then, in principle, nothing restricts us from doing the linearization around the point which the economy is currently at, rather than around the steady state.

By linearizing the model around the point which the economy is currently at, we minimize the approximation error in that period. But we can also do this sequentially, in each period - possibly resulting in lower approximation errors for the whole time range. Namely, because the linear solution lets us calculate endogenous variables recursively (i.e. depending only on predetermined variables and shocks), we can always update the coefficient matrices according to where the economy is currently. This way, by making linearization in each period, obtaining new coefficient matrices in each period and then calculating the values of the endogenous variables for the next period, we are essentially using a non-linear approach. That's why we can call this approach a Dynamic Linearization (DL).

Luckily, since the DL is recursive in nature, it lets us do all the calculations that are easily done on standard linear cases. This is well demonstrated below, both analytically and numerically. In particular, we show that the DL is almost as fast as linear solution methods, while it is almost as accurate as other more standard non-linear solution methods. However, we should note that this improvement in accuracy is only visible when the economy is far away from the steady state, i.e. when it is hit by a large (anticipated) shock or by a sequence of small shocks of the same direction or the model's propagation mechanism is persistent enough. At the same time, because of its recursive nature, we can, and below we do, easily adjust the standard Kalman filter to capture non-linear episodes in the US economic history. With this non-linear Kalman we can estimate unobservables better, at least in times of big changes in the macro-financial variables. In our view, this is probably the most important contribution of this approach.

The paper is organized as follows. The next section describes a simple non-linear dynamic model used for experimenting with the new solution methods. The following section then discusses the Combined Path approach in a nutshell and, using impulse response analyses and simulations, demonstrates its advantages. The fourth section describes the Dynamic Linearization algorithm and shows how it lets us do all of the analyses, done on linear models, in a non-linear fashion. The final section concludes.

2 A non-linear model for experimenting

In this section we describe a model economy that is used for comparing solution methods in terms of impulse responses, simulations and filters.[1] The model is essentially a standard, relatively simple, DSGE model (similar to Gali, 2007 or Smets and Wouters, 2007) but augmented with strong non-linearities. These non-linearities are both empirically as well as technically relevant: empirically in a sense that they feature the real world issues, experienced throughout the economic history; and technically in a sense that they are sufficiently strong to let us demonstrate the virtues of our proposed non-linear solution methods.

2.1 Model equations

The model consists of four main equations. Most of the structure is similar to basic New Keynesian DSGE models with dynamic IS curve, hybrid Phillips curve and Taylor-type interest rate reaction function. Let's first show more standard part of the model: the

[1] Doing comparison using more models is a task for future research

dynamic IS curve. This equation can trivially be derived from simple micro-foundations with intertemporally optimizing agents and habit formation:

$$y_t = \alpha_1 E_t(y_{t+1}) + \alpha_2 y_{t-1} - \alpha_3(i_t - E_t\pi_{t+1} - \rho) + \varepsilon_t^y \tag{1}$$

where y_t is the output gap in period t, E_t - an expectation operator with time-t information set, the parameters α_1 and α_2 depend on the degree of habit formation and the parameter α_3 is related to the risk-aversion in the household's utility function (not discussed here). In turn, i_t is nominal interest rate, π_t - inflation, ρ - steady state real interest rate, while ε_t^y is a preference (demand) shock. This equation does not feature non-linearities itself, but does propagate them to the rest of the variables both intratemporally as well as intertemporally.

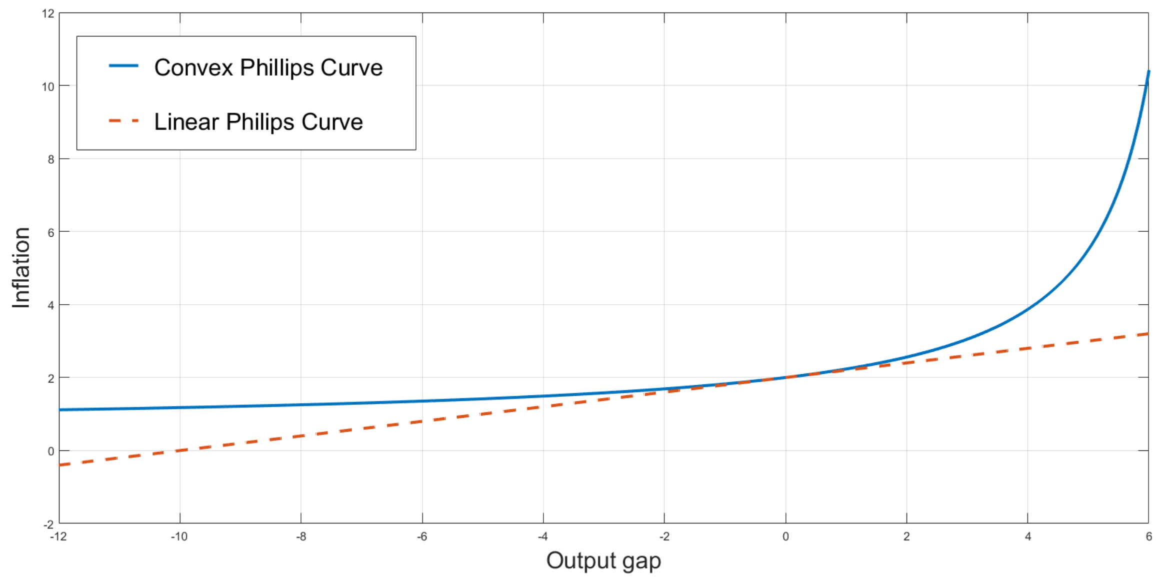
The other three of the four equations, in turn, are the direct sources of non-linearities. Starting with the hybrid Philips curve, apart from non-linear part, the equation is fairly standard - derived from micro-foundations by assuming firm's profit maximization with price stickiness and indexation to past inflation (for those who don't change their prices during a given period). As for the non-linear part, here we assume the convexity of the Phillips curve:

$$\pi_t = \beta_1 c_t E_t(\pi_{t+1}) + (1 - \beta_1 c_t)\pi_{t-1} + \beta_2 y^{max} \frac{y_t}{y^{max} - y_t} + \varepsilon_t^\pi \tag{2}$$

where $\beta_1 c_t$ captures the forward-looking part in the inflation process, which, apart from the parameter β_1, is affected by a non-linear process of credibility c_t (discussed below). ε_t^π at the end of the equation captures cost-push shocks. The convexity in output gap, in turn, is captured by the following component: $y^{max}\frac{y_t}{y^{max}-y_t}$, which makes the slope of the Phillips curve increase or decrease depending on the stage of the business cycle. In linear cases the slope would have been constant at β_2. Both of these two cases are illustrated by Figure 1, which depicts a contemporaneous relationship between inflation and the output gap for convex and linear Phillips curves.

This is both, interesting for technical analysis as well as policy relevant. The convexity

Figure 1: Non-linearity (convexity) in the Phillips Curve



Source: Authors' construction

makes the slope of the Phillips curve time-varying and state-contingent in a non-linear way. Solution method being able to efficiently handle this would be clearly advantageous. That's part of the reason why we consider it here. As for policy relevance, convexity seems to be prevalent in the real world too. In the US, for instance, after the Global Financial Crisis there has been increased acceptance of the fact that inflation does not react much to the output gap, while in pre-Volcker period it was clear that a very big increase in inflation (apart from oil prices) was largely brought about by a too high output gap. Both of these seemingly contradictory observations can be reconciled with the convex Phillips curve, where inflation reacts more to positive output gap and less to negative one[2]. In addition, as argued by Laxton et al (1998), it is still justified to assume the convexity in the Phillips curve and conduct monetary policy according to it even if we don't know whether the real world operates under convex, linear or even concave Phillips curve. In other words, for policy makers, it is less costly to err on the side of convexity than erring on the other. Laxton et al (1998) also show that the convexity of the Phillips curve easily explains the first-order benefits from pro-activeness in maintaining price stability.

[2]Another explanation is the credibility of the central bank, which we discuss below.

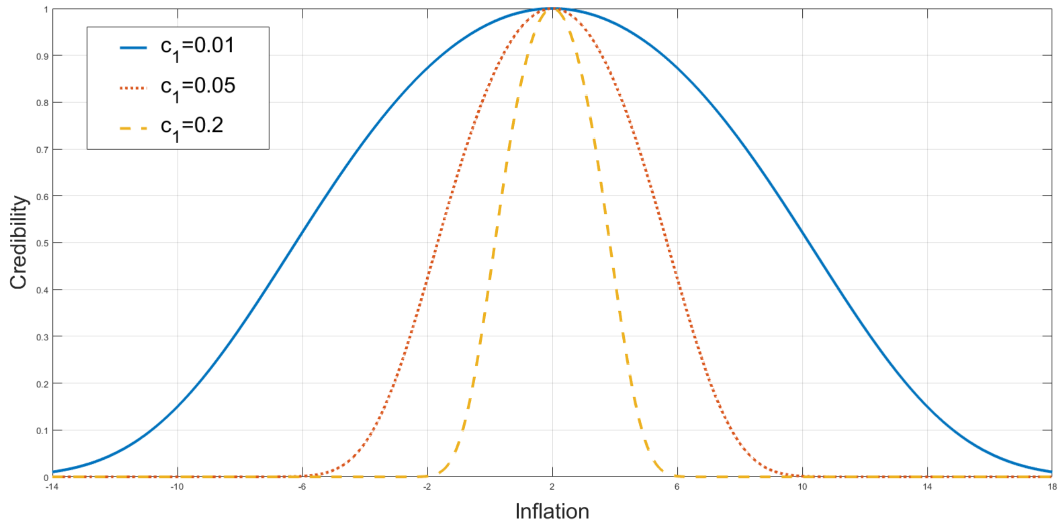
Next, as noted above, the degree of forward-lookingness in the inflation process depends on the variable c_t, which we call a monetary policy (or inflation target) credibility. In general, monetary policy credibility should be dependent on the success of a central bank in maintaining inflation close to its target. However, this dependence would be non-linear: it should not decline much when inflation deviates only a little from the target, while it should start to decline increasingly as inflation deviation from the target gets large. In addition, whenever the credibility declines from a maximum of one to a minimum of zero, afterwards, the public no longer cares about the size of this deviation, unless inflation returns to the target. Finally, the change in the credibility should not be abrupt, as it takes time to gain (accumulate) and, in many cases, lose the credibility. The following is the equation that takes all of these observations about the shape of the credibility process into account:

$$c_t = \rho_c c_{t-1} + (1 - \rho_c) 2 \big[1 - \mathcal{F}\big(c_1 \big[\pi_t - \pi^T\big]^2\big)\big] + \varepsilon_t^c \tag{3}$$

where ρ_c is the persistence of the credibility, $\mathcal{F}$ is the standard normal cumulative distribution function and $(\pi_t - \pi^T)$ is the deviation of inflation from its target. The parameter c_1 measures the sensitivity of the credibility impulse to inflation deviation from the target. In other words, the parameter c_1 could be interpreted as economic agents' collective tolerance towards central banks missing their inflation targets. The higher c_1 the more intolerant agents are. The shape of the credibility impulse (the non-linear part of the equation) for different values of c_1 is depicted on Figure 2. It can readily be observed that this impact is symmetric around the target. Credibility decreases at the same rate in both cases, whether inflation overshoots or undershoots the target to a sufficient degree. Benes et al (2017) and Tvalodze et al (2016) also model monetary policy credibility in this way, however their process is not symmetric around the target, while ours is.

If the credibility declines the inflation process becomes more backward-looking: with more agents indexing their prices to past inflation, the cost of price stabilization increases. In the extreme case of full backward-lookingness, the only way to decrease inflation would be to compress the output gap a lot (like Volcker did). On the other

Figure 2: Non-linear part of the monetary policy credibility process



Source: Authors' construction

hand, as the central bank returns inflation close to the target and maintains it there, the credibility increases and makes maintaining price stability much less costly.

Finally, the last equation of our model is the central banks' interest rate reaction function, however with effective (zero, in this case) lower bound being taken into account:

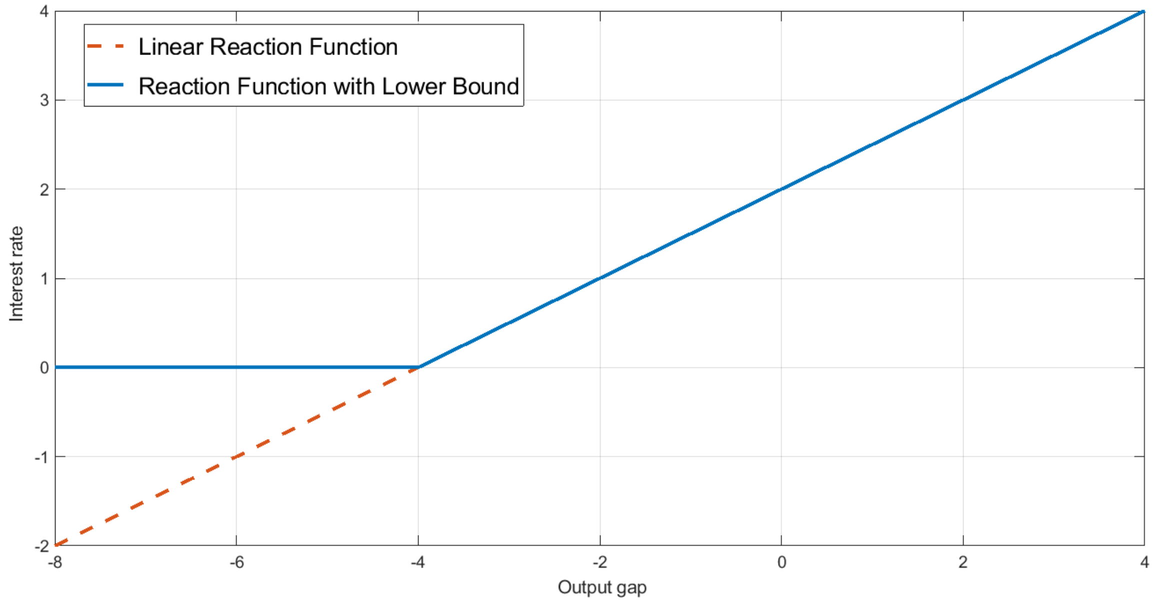
$$i_t = max\big(0, \rho_i i_{t-1} + (1-\rho_i)\big[\rho + \pi^T + \gamma_1\big(E_t(\pi_{t+1}) - \pi^T\big) + \gamma_2 y_t\big] + \varepsilon_t^i\big) \tag{4}$$

where ρ_i measures the degree of smoothing in the interest rate, while γ_1 and γ_2 measure the reaction of interest rate to inflation expectations and the output gap, respectively. The reaction function is modeled using the *max* function, in order to illustrate the ability of our solution methods to deal with kinks, and is displayed in Figure 3.

2.2 Calibration

Calibration is done based on standard DSGE models and empirical literature, where available. Namely, habit formation is assumed to be 2/3, so that $\alpha_1 = 0.6$ and $\alpha_2 = 0.4$. Risk aversion is set at unity (as is usually done), i.e. $\alpha_3 = 1$. On the other hand, the price stickiness is assumed and when prices do not change full backward indexation is used, while when they do they change approximately every 2-3 quarters on average,

Figure 3: Interest rate reaction to the output gap (with effective lower bound)



Source: Authors' construction

implying relatively flat Phillips curve, with the slope $\beta_2 = 0.2$ (in the neighborhood of the steady state). The degree of forward-lookingness in inflation process β_1 is set at 0.4, while the maximum level of output gap $y_{max} = 7$. The persistence of credibility process ρ_c is calibrated at 0.8 and the reaction of credibility impulse to the squared deviation of inflation from its target $c_1 = 0.01$. Interest rate smoothing is assumed by $\rho_i = 0.7$, while its reaction to inflation deviation and output gap are calibrated at $\gamma_1 = 1.5$ and $\gamma_2 = 0.5$, respectively. Inflation target is calibrated at 2% while real neutral interest rate is set at 0%, in line with some recent empirical estimates for advanced economies (Holston et al, 2017).

3 Combined Path

With the calibrated dynamic non-linear model at hand we can now describe the new solution methods and compare them to others. Until we move on to impulse response and simulation analyses, below we explain the basic idea behind the CP approach.

3.1 The basic idea

First, it is useful to recall what standard non-linear (stacked system) solution method does. Let's remember that a solution to a non-linear dynamic stochastic model

$$E_t\big[f(y_{t-1}, y_t, y_{t+1}, \varepsilon_t)\big] = 0 \tag{5}$$

is a set of values of endogenous variables $\{y_t, y_{t+1}, ...\}$ that satisfy equilibrium conditions in all periods, given values of exogenous variables. In equation 5, at any time-t period, y_{t-1} is given (predetermined), but in order to find y_t we also need y_{t+1}, which, in turn, itself depends on y_{t+2} and so on. If we assumed that y_{t+T} is close to zero (i.e. terminal condition), then we could solve the entire system for all t to $t+T$ periods at once. This is, in essence, a two-point boundary value problem and is solved by stacking the system (as discussed in Armstrong et al, 1999) or EP algorithm (proposed by Fair and Taylor, 1983).

Solving non-linear models of relatively large size with this approach and with sufficient accuracy becomes computationally very demanding, because the EP method tries to solve a non-linear system of equations of the size greatly bigger than the size of the model itself. This is so because one has to look at the number of periods sufficiently far ahead for the method to be accurate enough (i.e. terminal condition should be far into the future). If the model has N endogenous variables and we solve it for T periods, then EP has to solve a system of NT non-linear equations. In particular, for $f(\cdot)$ being the list of equilibrium conditions of the model (N equations), y_t - column of N endogenous variables and ε_t - that of exogenous variables (shocks), the system that EP tries to solve looks like the following:

$$F(x) = \begin{cases} f(y_{t-1}, y_t, y_{t+1}, \varepsilon_t) \\ f(y_t, y_{t+1}, y_{t+2}, \varepsilon_{t+1}) \\ ... \\ f(y_{t+T-2}, y_{t+T-1}, y_{t+T}, \varepsilon_{t+T-1}) \end{cases} \tag{6}$$

where $x = (y_t', y_{t+1}', ..., y_{y+T-1}')'$. Given y_{t-1} (initial condition) and y_{t+T} (terminal condition), $F(x)$ has NT number of unknowns. EP tries to find $x = x^*$ for which $F(x^*) = 0$. Now imagine a medium-scale dynamic model (similar to what many central banks or research institutions have) as big as with 50 equations (N) and the number of periods of 200 (T), then EP algorithm would have to solve a system of 10 000 non-linear equations - hardly a trivial task. For example, Newton algorithm, one of algorithms for handling this kind of task, will have to find (and several times) an inverse of Jacobian matrix that will be 10 thousand by 10 thousand. This takes much time, let alone convergence issues.

Now let's turn to the CP approach. The problem with the EP becomes complicated because in the model we have lags (predetermined variables) as well as leads (jump variables). However, if we were able to substitute leads with current variables, the algorithm would have to do much easier job. The idea is that, we can actually do this substitution by: (i) solving the model with first-order (or could be higher-order) approximation (ii) and expressing the time-$(t+1)$ variables in terms of time-t variables. In the case of linear approximation (*only* to expectations) the modified solution approach would have to solve only the following non-linear function (for time t):

$$f(y_{t-1}, y_t, \underbrace{Ay_t + B\varepsilon_{t+1}}_{\approx y_{t+1}}, \varepsilon_t) = 0 \tag{7}$$

where A and B are matrices obtained from linear solution methods (like Christiano, 2002 or Blanchard and Kahn, 1980). Now, for given y_{t-1}, this system has N, not NT, unknown endogenous variables and the same number of equations. This is the CP approach in a nutshell.

The advantages of this approach are: (1) it is computationally easier to solve a system of N equations T times as opposed to solving a system of NT equations at once; (2) in addition, in this case we don't necessarily have to solve the model for all T time periods, as we may be interested in just the first, say, 20 ($< T$) periods, while in the EP approach we are forced to solve for all T time periods. In other words, we do not have to make an assumption on when to end the simulation (which is a source of approximation error in the standard EP); (3) behaviorally, one could argue that it is also (maybe more) realistic to assume that model agents form expectations approximately (see Driscoll and Holden, 2014), i.e. they understand the probable future direction of the system, yet they don't fully understand how much would non-linearities change that outlook quantitatively.

Hence, this new solution approach seems to be making one reasonable assumption (and getting rid of another one used in EP), but results in much less computational burden. For demonstrating that the CP does indeed retain most of the accuracy of the EP, we can analyze impulse responses for the new approach and compare them to those of the standard EP as well as to linear solution methods.

3.2 Impulse response analysis

Let's compare impulse response functions (IRFs) across the three solution approaches (CP, EP and Linear) for the simple DSGE model described in Section 2. To begin with, consider the model with standard calibration described above and 1% negative output shock in the first period. The resulting impulse responses are depicted in Figure 7. As expected, our CP method and EP results are very close to each other, as well as to linear approach, thanks to the small value of the shock and only negligible non-linearities in the vicinity of the steady state. This confirms that the new MATLAB code for the CP is working well and, hence, we can keep on scrutinizing our approach further. Thus, let's check the robustness of the CP approach and demonstrate its advantages over linear and/or EP methods by considering various policy-relevant scenarios and across different calibrations of the model.

Consider first a situation where an economy is hit by a big negative output shock that pushes the output gap down to -10% for the first two quarters (Figure 8). In other words, imagine a situation where policy-makers ask what would be the macro implications if such a thing happens to the output gap. As we see from the figure, differently from the previous case with only -1% output shock, IRFs for the three solution approaches do not move close to each other anymore. While EP and CP almost coincide, linear method diverges significantly. This could be explained by the fact that the approximation around the steady state (used by the linear method) is not precise anymore due to a large output shock and now-activated non-linearities. The fact that our method produces the same results as the EP / perfect foresight stacked system approach used by Dynare proves that our method is robust to significant deviations from the steady state and still produces accurate results.

The most important thing here that highlights the advantages of the CP approach over linear one is how accurately it captures the zero lower bound episode (like the EP method). As seen from the interest rate subplot of Figure 8, interest rate drops to almost -4 % in the linear case (which, of course, is unrealistic) and makes recovery fast, while the CP (as well as the EP) takes into account the lower bound restriction (an empirical observation that we know exists in real world) and complicates the process of returning the economy back to the steady-state. Another important observation is the level of inflation, which in linear case reduces significantly in the beginning, while overshooting the target later due to aggressive monetary accommodation. In contrast, when we take non-linearities into account, using our CP approach (as well as the EP), we see that inflation exhibits weaker downward response and goes back to target only slowly. Two non-linearities contribute to this. First, as credibility worsens (due to moving off-target), inflation process becomes more rigid. Second, due to convexity in the Phillips curve, negative output gap does not cause disinflation as much as positive output gap would cause inflation. These further underlines our proposed method's precision in capturing non-linear phenomena, comparable to Dynare's non-linear (perfect foresight) method.

Speaking about deep recessions, interesting policy relevant question would also be to ask what would happen to the real economy if additional negative shock happens even when we are already at the zero lower bound. This is exactly the scenario shown in Figure 9. This figure is comparable to Figure 8 in a sense that the level of shock is the same in the first period on both figures, however on Figure 9 additional one percentage point negative output shock hits the economy in period two. If we observe the plots closely we see that the linear method suffers the same deficiencies of bearing large approximation errors (most notably missing the zero lower bound). As for the EP method it stays on zero lower bound for shorter time span than was the case on Figure 8. Intuitively, this means that additional negative output shock stimulates the economy and pushes it back to steady state, which is a clear contradiction to economic intuition. This probably has to do with technical problems in the non-linear solution algorithm of Dynare. As for our CP approach, the IRFs show how difficult such a situation would be and that we would need to stay at the zero lower bound for much longer - something one would expect from economic intuition and our prior experience with the global financial or eurozone crises.

Now let's compare the solution methods by checking their robustness against different calibrations of the model (but with the same size of the shocks as in the case of Figure 8). For this, we consider three different cases for the persistence parameters α_1 and α_2 in the dynamic IS equation. This is the question policy-makers may also ask: what would happen if we are hit by such big output shock, as discussed in the previous case, but the economy has more/less rigid propagation mechanism. For each of the three solution methods we consider three scenarios: 1. the IS equation is more forward-looking ($\alpha_1 = 0.6$ and $\alpha_2 = 0.4$); 2. the IS equation puts the same weights on forward-looking as well as on backward-looking components ($\alpha_1 = 0.5$ and $\alpha_2 = 0.5$); and 3. the IS curve is more backward-looking ($\alpha_1 = 0.4$ and $\alpha_2 = 0.6$). First, let's consider the linear solution (Figure 10). The size of the negative shock on output is the same as before, so that when the model is more forward-looking output drops to -10% as before (baseline calibration). However, as the persistence of household's and firm's behavior increases output gap deepens further (as would be expected intuitively) in the subsequent periods. The response of the interest rate and inflation is proportional. The more backward-looking the model becomes the lower the required interest rate to stimulate the economy and now-lower inflation.

Non-existence of the lower bound for the interest rate is evidently still the biggest drawback of the linear approach. This is not the case for the EP method. However, it suffers from another problem. Figure 11 shows the IRFs as calculated by the EP method across the three calibrations. The blue path corresponds to less persistent version of the model and the results are similar as described before (baseline calibration). However, the more backward-looking the model is the smaller and less persistent the output gap becomes. EP approach still captures the zero lower bound for interest rates. But the fact that more backward-looking model calibration produces less persistent output gap dynamics is inconsistent with economic intuition and highlights the limitation of the EP approach[3]. Our novel approach, however, does take the zero lower bound into account and also does not provide dynamics inconsistent with economic intuition (Figure 12). As it is clear from Figure 12, more backward-looking version of the model induces stronger persistence in all three variables while the length of a spell at the zero lower bound increases significantly. This (more backward-lookingness) could be a useful explanation of why some advanced economies remained at the lower bound for so long.

The fact that the CP is more robust than the EP is further underlined if we now consider positive output scenario (output gap increasing to 4%) but across different interest rate reaction functions. This is shown on Figures 13, 14, 15, where we show solutions for the different calibrations using linear, EP and CP approaches, respectively. In all three cases we applied the same level of positive output shocks. Different calibrations measure how strongly the central bank reacts to inflation. Intuitively, as interest rate reaction function becomes more reactive (equivalently, as γ_1 increases) we should expect to observe smaller output gap and smaller deviation of inflation from the target level. This is well shown by the linear approach; however, it fails to incorporate the quantitative part correctly, due to missing important non-linearities. On the other hand, Dynare's non-linear perfect foresight solver is supposed to take these non-linearities into account. Here too, though, in many cases it fails produce sensible dynamics. For instance, in this scenario if we look at the inflation subplot on Figure 14, we see that for a *less* reactive

[3]This is when the Dynare does provide the results. In many cases, it cannot even solve and crashes.

central bank as output gap becomes positive inflation deviates downward altogether, contradicting to what economic intuition would suggest. Again, this is probably a computational or algorithm-related issue in Dynare.

This problem does not seem to arise when we solve the system using our CP method. Like linear approach this method captures the economic intuition that a more reactive central bank would deliver less macroeconomic volatility. It also captures that there would be first-order benefits from being proactive in maintaining macroeconomic stability, something that linear models fail to capture (so called certainty equivalence). In other words, the CP approach does not face as much technical difficulties in the solution algorithm (which is probably the reason why the EP approach fails) and does not miss important non-linearities altogether (which is the reason why linear approach is inaccurate).

Another example to show why CP overcomes the other two approaches is dealing with embedded non-linearities in the Phillips curve. As described in section 2, Phillips curve in our model is strictly convex, making the slope of the curve dependent on the state of the economy. In particular, the slope is an increasing function of the output gap. In other words, if the economy is expanding an additional increase in output gap will produce higher additional deviation from inflation target, while if the economy is in a recession an additional deepening of the negative output gap will reduce inflation marginally less. These effects are captured by both the EP and CP but not by linear approach (see Figures 16 and 17). Figure 16 depicts 2% positive output shock in the first and the second periods each while Figure 17 represents symmetrical negative output shock (for the first two periods).

There are differences in how inflation behaves responding to two consecutive output shocks. When reacting to the positive output shock in the second period when the economy is already in a boom the CP captures the increasing slope of the Philips curve (subplot of inflation on Figure 16) producing higher inflation than in the linear approach. The situation is symmetrical in a sense when considering two consecutive negative output shocks. Linear method results in much lower inflation than CP (plot for inflation on Figure 17). It is true that by looking at Figures 16 and 17 one might think that the EP is better in capturing the convex nature of the Philips curve; however increasing the size of the shock (even marginally) takes down the robustness of the EP. In particular, Figure 18 describes the case when the economy is hit with an additional percentage point of shock in each of the first two periods (i.e. additional 2pp in total). While both CP and linear methods still produce economically intuitive results, the EP depicts very strange outcome - for positive output shocks the economy goes into a deep recession, inflation falls and interest rates decline down to the zero lower bound. This seems to be a technical complication in the standard routine of Dynare. On the other hand, the CP algorithm still handles this scenario well while taking the non-linearities into account - making it superior to both standard linear and non-linear approaches used by Dynare.

Finally, another advantage of the CP approach deals with time and computational constraints. One of the major drawbacks of using standard non-linear solution methods with large scale DSGE models is the huge time resources needed. Even on computers with high computational powers non-linear methods like stochastic Extended Path could take many hours. This feature of non-linear solution methods makes them very inconvenient in empirical and policy-making contexts. Hence, finding alternative solution methods that deliver results faster at almost no cost in accuracy is clearly advantageous. Our proposed CP approach seems to achieve just this. We have done a simulation in response to random shocks in every period over 1000 periods using both EP as well CP approaches. As it was clearly visible in the simulation outcome, both approaches produced practically identical results. However, simulating the baseline model using the CP required six times less time. Namely, the CP approach ran the simulation in about 14 seconds while the EP took about 94 seconds[4].

[4]Standard deviations used in generating the (pseudo) random numbers for the model shocks were assigned arbitrarily - we just started from big values and kept reducing standard deviations until Dynare's '*extended_path*' command didn't crash and could produce results. As expected from the IRFs above, our CP approach didn't have that kind of problem.

4 Dynamic Linearization

The DL approach to solving non-linear dynamic models exploits two features of the linear solution methods: ease of being used for different kinds of exercises and their recursive nature. With this, as discussed below, the final outcome should be complexity of only linear methods, but accuracy of non-linear ones.

4.1 The basic idea

The idea behind the DL on the most basic level is to make first-order (or higher-order) approximation not around the steady state of the model, but around the current stance[5] of the variables and to do so in each period. Usually, when the linearization is performed (around steady states), one gets the following approximation:

$$\begin{aligned} E_t\big[f(y_{t-1}, y_t, y_{t+1}, \varepsilon_t)\big] \approx \quad & f(\bar{y}, \bar{y}, \bar{y}, \bar{\varepsilon}) + f'_{y_{t-1}}(\bar{y}, \bar{y}, \bar{y}, \bar{\varepsilon})(y_{t-1} - \bar{y}) \quad + \\ & f'_{y_t}(\bar{y}, \bar{y}, \bar{y}, \bar{\varepsilon})(y_t - \bar{y}) + f'_{y_{t+1}}(\bar{y}, \bar{y}, \bar{y}, \bar{\varepsilon})\big(E_t\big[y_{t+1}\big] - \bar{y}\big) \quad + \\ & f'_{\varepsilon_t}(\bar{y}, \bar{y}, \bar{y}, \bar{\varepsilon})(\varepsilon_t - \bar{\varepsilon}) \end{aligned} \tag{8}$$

where f'_x denotes the first-order derivative of the model equations $f(\cdot)$ with respect to x and bars over the variables represent their reference points (in this case, steady states).

The problem with this kind of approximation is that it gets more and more inaccurate as we move away from steady states. For instance, if today the system stands somewhere far from steady state and some new shock materializes, then the effect of this shock would be severely incorrectly approximated. Yet these kind of questions arise often both

[5]Even though we came up with the approaches and did our work completely independently, it just came to our attention that Evans and Phillips (2015) proposed an approach similar to our second method presented here (that we dubbed as Dynamic Linearization). But since we found out about this long after our work was finished and our paper was about to be published, we decided to discuss the differences only briefly. Namely, they follow a slightly different computational approach, even though the general idea is the same. In addition, and more importantly, they do not discuss the extension of the Kalman filter that results in a non-linear filter. Here we show that having a non-linear version of the Kalman filter could make it possible to analyze economic phenomena especially important for policy-makers throughout the world.

at policy-making as well as research institutions, especially after the Global Financial Crisis. One way to get around this problem and reduce approximation error is to make a first-order approximation around the *current stance* of the endogenous variables and *not* around the steady states. Importantly, we do this in each period.

With this idea in mind we can now describe the DL approach in more details. In particular, we have to:

1. In period 1 linearize the model around the initial conditions (usually steady states) and obtain the linear solution of the form $y_t = A_1 y_{t-1} + B_1 \varepsilon_t$, where A_1 and B_1 are (reduced-form) solution matrices for this particular approximation (period). y_0 (initial condition) is given. This way we get y_1;

2. For every next period ($t \geq 2$) linearize the model around the values of the variables predicted by the previous period solution, i.e. around $\bar{y}_{t-1} = y_{t-1}$ (this is predetermined), $\bar{y}_t = A_{t-1}y_{t-1} + B_{t-1}E_{t-1}(\varepsilon_t)$, $\bar{y}_{t+1} = A_{t-1}^2 y_{t-1} + A_{t-1}B_{t-1}E_{t-1}(\varepsilon_t) + B_{t-1}E_{t-1}(\varepsilon_{t+1})$ and $\bar{\varepsilon}_t = E_{t-1}(\varepsilon_t)$, where all of those (bar variables around which we linearize) are known as of time-t, i.e. they are just numbers;

3. Given the linearized model from step 2 solve it using linear solution methods to obtain new matrices A_t and B_t and the values of y_t. Go to step 2 if the next period values of the endogenous variables are needed.

Analytically, for every period t:

$$
\begin{aligned}
E_t\big[f(y_{t-1}, y_t, y_{t+1}, \varepsilon_t)\big] \approx \; & f(\bar{y}_{t-1}, \bar{y}_t, \bar{y}_{t+1}, \bar{\varepsilon}_t) \quad + \\
& f'_{y_{t-1}}(\bar{y}_{t-1}, \bar{y}_t, \bar{y}_{t+1}, \bar{\varepsilon}_t)(y_{t-1} - \bar{y}_{t-1}) \quad + \\
& f'_{y_t}(\bar{y}_{t-1}, \bar{y}_t, \bar{y}_{t+1}, \bar{\varepsilon}_t)(y_t - \bar{y}_t) \quad + \\
& f'_{y_{t+1}}(\bar{y}_{t-1}, \bar{y}_t, \bar{y}_{t+1}, \bar{\varepsilon}_t)\big(E_t[y_{t+1}] - \bar{y}_{t+1}\big) \quad + \\
& f'_{\varepsilon_t}(\bar{y}_{t-1}, \bar{y}_t, \bar{y}_{t+1}, \bar{\varepsilon}_t)(\varepsilon_t - \bar{\varepsilon}_t)
\end{aligned} \tag{9}
$$

where, on the right hand side, all the inputs to $f(\cdot)$ and to its derivatives with respect to each of all variables are known as of time t (and given above in step 2). Hence, for solution purposes at time t this approximation is linear. Yet, at the same time, because we update the coefficient matrices and obtain this kind of linear solution in every period, overall it gives us the accuracy of non-linear solution method.

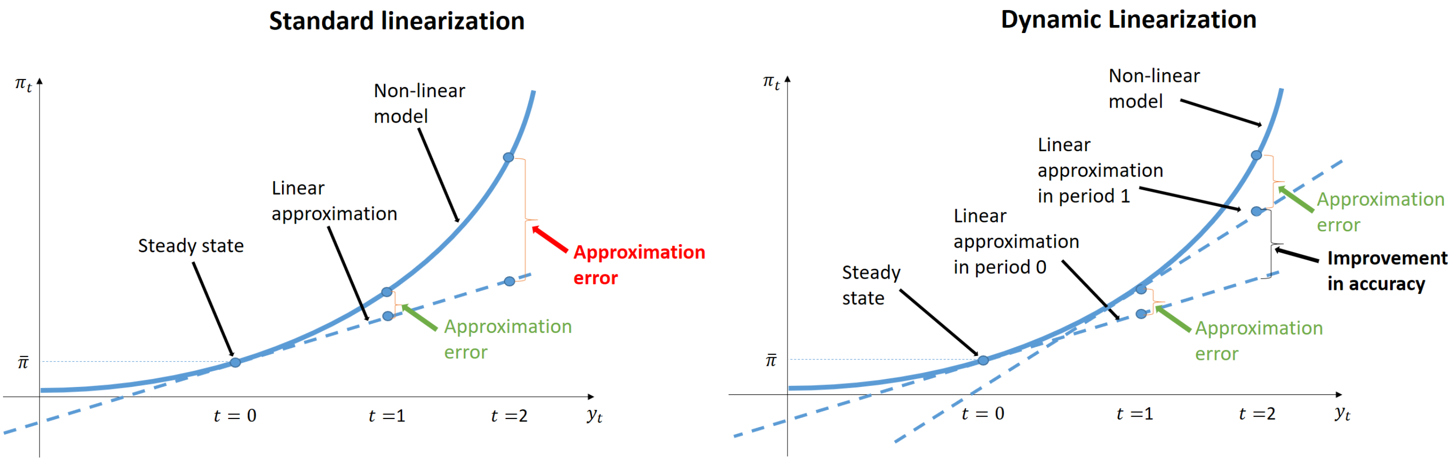
The advantages of this approach are: (1) it is computationally much easier to do, relative to other well-known non-linear solution methods. This is so because we just have to *evaluate* the functions $f(\cdot)$ and $f_x^{'}(\cdot)$ used above (functional forms of which are usually available from model builders) at particular values of inputs, i.e. no optimization algorithm is involved; (2) its accuracy could be much greater than standard linear solution methods when a system experiences a sequence of shocks. However, the approach may also give more accurate picture even in standard impulse response analyses when only one shock is used once in period 1. The reason for this is that since the solution coefficient matrices change over time, it could make model propagation mechanism more (less) persistent over time, depending on the non-linear features of the original model; (3) it could easily be used in standard Kalman filtering or Bayesian estimation (the latter not being discussed here), e.g. as opposed to much more computationally demanding particle filter.

The difference between the standard linear approach and the DL is illustrated on Figure 4, where one can clearly see where the improvement in the accuracy for the DL comes from - never letting the approximation fall behind the recent dynamics in the system. Finally, for demonstrating the benefits of the DL numerically, we can do the impulse response or Kalman filtering exercises on our non-linear DSGE model described in Section 2.

4.2 Impulse response analysis

As in the case of the CP approach, let's first start by comparing the IRFs derived using the DL and standard linear approaches for a small shock size. The expectation is that both approaches should give very similar results given the small size of the shock and starting point being the steady state. The results of this exercise are shown on Figure 19, which does indeed confirm that both approaches produce virtually identical results.

Figure 4: Difference between standard linearization and the DL (simple example)



Source: Authors' construction

These calculations also show that our MATLAB code for the DL is working well and we can continue scrutinizing further.

Next, we consider the case when the shock size is large enough to make the effective (zero in this case) lower bound binding. The results are depicted in Figure 20. This exercise is essentially the same as the one performed for CP approach on Figure 8. As we already noted in subsection 3.2 linear method reveals a number of deficiencies: (1) output gap reverts back to the steady state too fast; (2) this being due to missing the zero lower bound constraint (i.e. interest rate going deeply into negative territory and providing huge accommodation); and (3) inflation reacting too much to the negative output gap. The DL solution approach alleviates all of these problems to some extent. Namely, even though the DL cannot capture the interest rate lower bound restriction *until* interest rates actually go into negative territory[6], afterwards it takes this lower bound constraint into account - leading to acknowledging lower level of accommodation, slower reversion of the output gap back to steady state and lower reaction of inflation (due to capturing the convex nature of the Phillips curve[7]).

[6]This is due to the fact the solution matrices are calculated based on previous period values of endogenous variables. When interest rates become negative only then is the derivative of the interest rate reaction function (evaluated at that negative rate) equal to zero.

[7]Similarly to the interest rate reaction function, the convexity of the Phillips curve is captured by the DL only when the model has already been gone off the steady state.

Now, let's change the parametrization of the model and check how DL performs in relation to the pure linear method. This exercise is closely related to the Figures 10, 11 and 12 in subsection 3.2. To remind the reader, we change the values of the parameters in the dynamic IS equation (1). In particular, we assign the following values: $\alpha_1 = 0.4$ and $\alpha_2 = 0.6$. This way, the model becomes more backward-looking and the expected results should be higher persistence of the system and more difficulty in providing the sufficient amount of accommodation. This is indeed what we observe on Figure 21 for the DL approach, but, clearly, not for the linear one. In case of solving the model with the DL, interest rate remains on effective lower bound for almost 15 quarters (nearly 4 years), while output as well as inflation take approximately the same time to recover to the equilibrium level. The linear approach, on the other hand, falls behind this economic intuition significantly as the recovery is much faster than one with the memory of the Great Recession would expect, accompanied with so much unrealistically negative interest rates and huge accommodation that pulls the output gap back to its steady state quickly. This is clearly at odds with the reality and such a big approximation error stems from the fact that the linear solution approach neglects what happens off the vicinity of the steady state.

While speaking about how the DL reduces the approximation error present in linear method, it is worth mentioning that the significant improvement in the accuracy of the DL is mostly present when we have an anticipation that the system will not go back to the steady state quickly (i.e. when the shock is very big or anticipated, or there is a sequence of many, possibly unanticipated, shocks of the same direction in a row). Figure 22 shows impulse responses for positive output shock for the case when the shock in the second period is anticipated, while Figure 23 depicts the same scenario but with the shock in the second period being unanticipated. Indeed, if we compare the DL and linear solution methods on those two figures we see that the difference between the two methods is more pronounced when the shock is anticipated. Since the DL updates solution coefficient matrices according to where it expects the system to be in the next period, these (reduced-form) solution matrices induce much less approximation error when we do incorporate the expectation that the system will remain off the steady state (i.e. anticipate).

This does not mean, though, that anytime unanticipated shock hits the model, the DL would have the same approximation error as the linear approach. It just means that, for the DL to be much more useful than the standard linear solution method, you either need to be dealing with anticipated shocks or dealing with the economy persistently being off the steady state (due to anything). The latter cases are, indeed, prevalent in real world and can most effectively be captured by bigger macro-financial models. Applying the DL approach to that kind of richer models is a task of future research.

To sum up the results of impulse response analysis, we can see that the DL may not have such a big accuracy improvement when we are talking about small or short-duration deviations off the steady state. However, when there are large shocks or we stay away from the steady state for sufficiently long, as in real world, then the DL, by incorporating non-linearities into the solution algorithm, provides significant improvement in accuracy. In case of our simple model, this is best seen by simulating a big negative output gap shock that makes the interest rate lower bound binding. In short, the DL truly bears the virtue of being much more accurate than the linear method in most of the realistic cases, while maintaining almost the same level of computational complexity.[8]

4.3 Non-linear Kalman filter

Despite encouraging outcomes from the IRFs in the previous subsection, in our view, the most important contribution of the new DL approach is its ability to bring the non-linear model to data very easily. Indeed, as already mentioned, filtering non-linear models, which is very important for these models to be actively used in practice, has been difficult with other solution methods. Solving this problem even partly, something

[8] It is also worth noting that we did a simulation with randomly generated shocks using both linear and DL approaches over sufficiently many periods. The linear approach, as expected, shows that positive shocks cancel out negative ones so that throughout the whole simulation range, on average, inflation is on target and output gap is zero. This is not so in the case of the DL solution approach - here, on average, inflation is slightly above target, while output gap is slightly negative. This is consistent with the idea of convex Phillips curve and proactiveness in maintaining price stability being beneficial (see Laxton et al, 1998).

our DL approach seems to be doing, is a big step forward. Standard Kalman filter works by evaluating the likelihood of the state-space system of the following form:

$$x_t = Ax_{t-1} + \varepsilon_t, \qquad x_t \in \mathbb{R}^n,\ A \in \mathbb{R}^{n \times n}, \qquad \varepsilon_t \sim \mathcal{E}_t = \mathcal{N}(0, Q) \tag{10}$$

$$y_t = Fx_t + v_t, \qquad y_t \in \mathbb{R}^m,\ F \in \mathbb{R}^{m \times n}, \qquad v_t \sim V_t = \mathcal{N}(0, R) \tag{11}$$

where x_t is an n-vector of state (endogenous) variables with its vector of shocks ε_t and y_t is an m-vector of space (measurement) variables with its vector of shocks v_t. A is a coefficients (solution) matrix that describes the dynamics of the system, while F links the endogenous and measurement variables. The covariance matrix Q depends on the standard deviations of the model (structural) shocks and model parameters, while the covariance matrix R defines the standard deviation of measurement errors.

The problem with this standard approach is that it involves linearizing the original model around the steady state (to calculate the matrix A) and, hence, doesn't take the non-linearities into account. For instance, if inflation data has been hovering persistently above the target level when output is close to its potential level, standard Kalman filter will decide that we are very unlucky in that we are persistently experiencing a sequence of positive cost-push shocks in every period. Even if such a case would be very unlikely (cost-push shocks usually happen in both directions, at least throughout sufficiently long horizon), standard Kalman filter would have no other way to explain the data. However, from the original non-linear model we know that, whenever the credibility of the central bank erodes, inflation process becomes unanchored. And with unanchored inflation expectations actual inflation could very well hover persistently above the target level, even if we are not experiencing any cost-push shock.

The problem there is that the credibility process in the model is non-linear and, hence, standard Kalman filter doesn't even know about the existence of this process. This stems from the following fact: linearizing the credibility equation and the Phillips curve around the steady state yields

$$
\begin{aligned}
c_t = &\quad \rho_c c_{t-1} + (1-\rho_c)1 + \varepsilon_t^c && (12)\\
\pi_t = &\quad \beta_1 E_t(\pi_{t+1}) + (1-\beta_1)\pi_{t-1} + \beta_2 y_t + \varepsilon_t^\pi && (13)
\end{aligned}
$$

and since the credibility variable doesn't show up in the linearized Phillips curve, the most likely estimate of the credibility shock by the Kalman filter would be zero[9], i.e. $c_t = 1, \forall t$. This is so because there's no need to make the value of the credibility variable deviate from the steady state - after all, it has no effect on the system (when linearized around the steady state) at all.

The way our DL approach partially solves this problem, brought generally by the existence of non-linearities, is based on the fact that Kalman filter is recursive in nature. The (reduced-form) solution matrix above, A, can very well be made time-varying. But the whole idea of the DL isn't only that it can be made time-varying, but also that it can be made dependent on the stance of the economy, in line with the logic discussed in subsection 4.1. Namely, whenever the economy diverges from the steady state, we make new linearization around the point where the economy is currently at and, after solving this newly linearized model, we obtain new (reduced form) solution matrix that can be used for calculating the values of the endogenous variables for the next period. In other words A becomes a non-linear function of the endogenous variables, i.e. $A_t = A(x_t)$.[10] In other words, the state part of the state-space representation would become the following:

$$
x_t = \underbrace{A_{t-1}}_{\approx A(x_{t-1})} x_{t-1} + \varepsilon_t \qquad (14)
$$

Since, in equation (14), we know the value of the matrix A_{t-1} in period t (because

[9]This helps maximizing likelihood function, since these shocks are distributed normally around zero.

[10]This will be a non-linear function the shape of which would directly depend on the non-linearities in the original model.

we know x_{t-1}), we can still do the Kalman filtering without any problem. What's important here is that our matrix A changes in every period in such a way that reduces the approximation error of the non-linear model. In case of the example above, when inflation deviates from the target level, from the next period the credibility will start to decline[11] - making inflation process more backward-looking. This backward-lookingness would mean that the coefficient that links the actual inflation to the previous period inflation will start converging towards one. By this being reflected in the newly calculated matrix A, we would get inflation remaining at the above-the-target level for much longer, even if there is no other shock hitting the economy. By using the shocks less frequently and explaining the data with the endogenous part of the model, we obtain much better fit in terms of the likelihood function - interpreting the data in a more convincing and intuitive way.

With this approach at hand we can now apply the procedure to a real world example and compare the results with the standard Kalman filter. For this exercise we use the US data on real GDP, inflation and federal funds rate for the period 1955Q1-2019Q1.[12] During filtering we assigned the standard deviations of the Phillips curve (cost-push) shock, the IS equation (output gap) shock and reaction function (interest rate) shock to one. As for the standard deviation of the credibility shock, we assigned to it a negligible number in order to make the system estimate the credibility endogenously. In addition, we used inflation and interest rate data in the model directly, but detrended the real GDP data by the HP filter to come up with the output gap values over the history. For this reason we assigned a positive value to the standard deviation of the measurement error only for the output gap.[13]

With this data, model and calibration we get the estimates of the unobservables - values

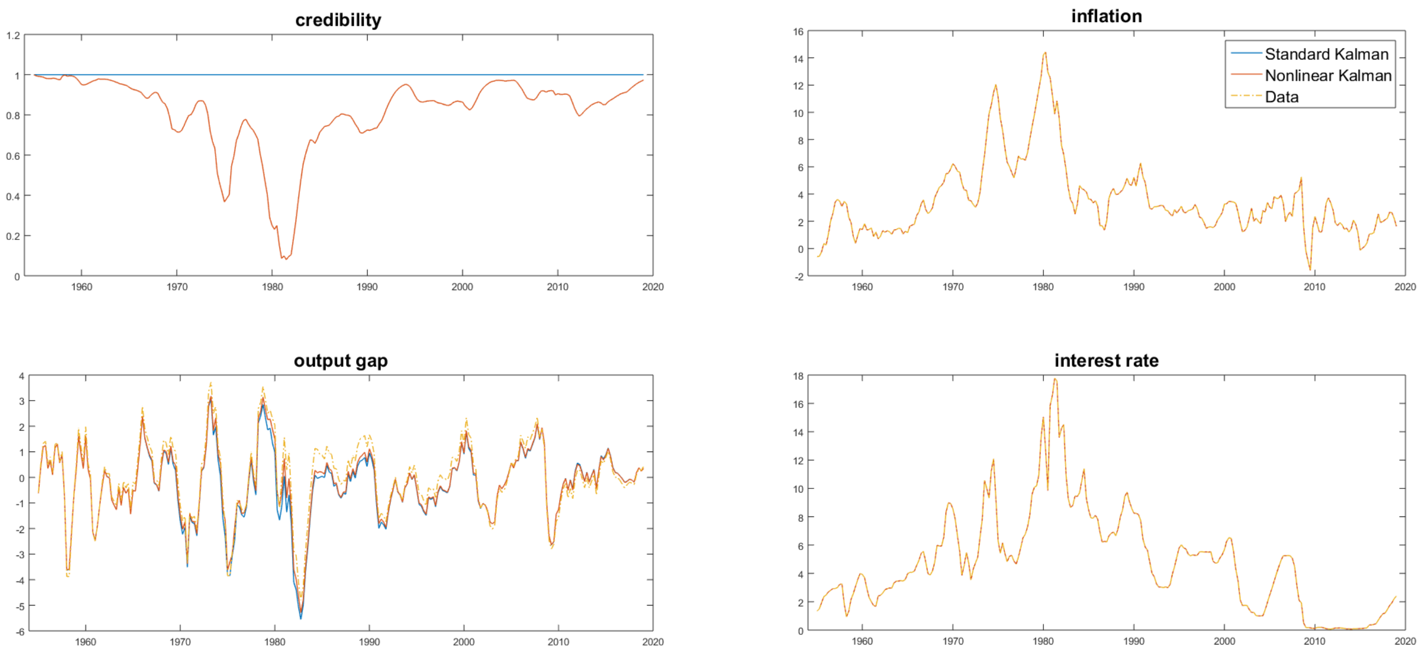
[11]If we linearize the credibility equation and the Phillips curve around the points different from the steady state, then the Phillips curve (linearized this way) would start to depend on the value of credibility, while the credibility equation would be the following: $c_t = \rho_c c_{t-1} + (1 - \rho_c) g(\bar{\pi}_t) + \varepsilon_t^c$, where $\bar{\pi}_t$ is the reference value around which we linearize (can always be different from the steady state) and $g(\cdot)$ is a function that directly derives from the derivative of the original credibility equation with respect to inflation and is from 0 to 1.

[12]The data has been obtain from FRED, Federal Reserve Bank of St. Louis.

[13]Meaning that R matrix above would have all zeros expect the place on its' diagonal which goes to the output gap measurement equation.

of all the shocks and the values of the credibility variable over the history. If we do this by the standard linear filter we would get an estimate of the credibility always equaling one. But arguing that the credibility doesn't change endogenously clearly is not true; it is only the result of the standard Kalman filter not being able to capture this non-linear process. However, combining our DL approach with the Kalman filter - which we call a non-linear Kalman here - doesn't suffer from the same problem. Namely, whenever the first-order approximation of the credibility equation is done around the point different from the steady state, the value of the credibility becomes dependent on the history of inflation and its deviation from the target - as in real world and our original model. Since the resulting non-linear Kalman takes this into account, this time we get a time-varying estimate of the credibility over the history.

Figure 5: Kalman filter results for endogenous variables (linear and non-linear)



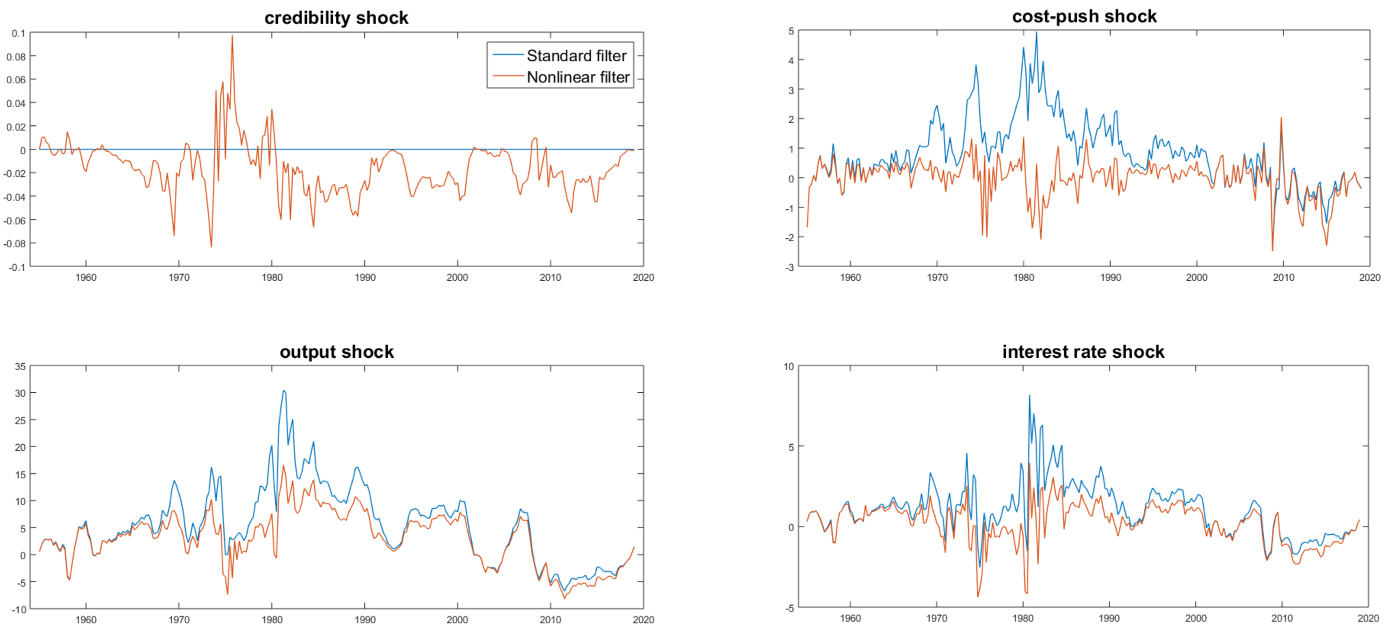
Source: Authors' calculations

Looking at Figure 5[14] we see exactly as discussed above. The estimate of the credibility using the standard (linear) Kalman is always full (one), while it is time-varying when

[14]The output gap differs slightly for 'data', linear estimate and non-linear estimate because, as discussed above, there is a measurement error that we incorporated into the output gap measurement equation.

estimated using our new approach (non-linear Kalman). As expected, our non-linear Kalman estimates the credibility to be significantly less than one during periods of persistently off-target inflation (mostly higher than target). This is especially evident during 1970s and early 1980s and can be attributed to excessively easy monetary policy at that time which led to unanchoring of inflation expectations. This has been reversed after the start of the Volcker era when monetary policy was tightened significantly, output gap went into negative territory and inflation started to decline towards the level consistent with the notion of price stability (which in the model is taken as 2% inflation). After that the credibility had its modest ups and downs but overall it seems to have remained close to one. This story, estimated by the non-linear Kalman, is in line with our intuition - we knew credibility was compromised back then, but had restored since Volcker era. We just couldn't show it with the standard linear Kalman filter. With our new approach - we can.

Figure 6: Kalman filter results for structural shocks (linear and non-linear)



Source: Authors' calculations

The advantage of the non-linear Kalman is also starkly visible if we look at the estimates of our structural shocks (Figure 6). First, let's have a look at the estimates provided by the standard linear Kalman. Here, as expected, the inflation rate during 1970s and early 1980s is explained by (unrealistically) big and many cost-push shocks in a row. This story would argue that we were just unlucky when we had that big inflation in that period. A similar (unlikely) story goes with the other shocks. But this clearly isn't very convincing. This linear estimates are missing non-linearities which we believe would had been activated at some points in economic history (e.g. during 1970s as regards the credibility). This is confirmed by our non-linear Kalman, which shows that it was not really a bad luck. Persistently high inflation was more a story of central bank's reduced credibility, which (and by implication inflation as well) was then reversed again back to normal levels by more responsible monetary policy. Comparing the estimates of the shocks, it is clear how huge would be differences in terms of likelihood function. Linear Kalman was forced to use a very unrealistic sequence of shocks (greatly reducing the likelihood), while our non-linear Kalman (thanks to the non-linear endogenous part) was able to explain data with the values of shocks that was, on average, very close to zero. Being able to explain history with shocks close to zero maximizes the likelihood. In other words, the story provided by the non-linear Kalman is clearly much more convincing than that of the linear one not only intuitively, but also econometrically. For future research, filtering the richer set of data with richer DSGE models is certainly a very interesting exercise.

Finally, it is worth noting that such non-linear estimates can be obtained by other non-linear filters as well, like particle filter. But they take so much time that it makes them unfeasible in everyday work of economists (see Schorfheide, 2016). As for our non-linear Kalman filter it takes practically the same amount of time to run as the standard Kalman filter, i.e. we are talking about (mili)seconds. This is what makes the DL approach to Kalman filtering attractive for researchers interested in non-linear models.

5 Conclusion

The paper proposes two new solution methods, the Combined Path and Dynamic Linearization, which try to contribute to the literature by providing more efficient and/or accurate solution approaches when dealing with strongly non-linear dynamic models. The CP approach, in a nutshell, substitutes jump (forward-looking) variables with something that both retains many features of forward-lookingness but simplifies the solution problem. This is done by combining linear solution methods with the usual non-linear (stacked system) algorithm *only* when it comes to jump variables. The impulse response analysis and simulations on non-linear models do show that this approach is an improvement, at least for the particular (strongly-enough) non-linear model we considered. Namely, the accuracy is almost the same as fully non-linear approach, while computer and time resources needed are about six times less.

On the other hand, the DL approach relies on the facts that non-linear models can be linearized around (nearly) any point we like, not necessarily around the steady state, and that linear solutions to dynamic models are recursive in nature. Taking these features into account we can essentially do a non-linear analysis using linear apparatus in a recursive fashion: making linearization in each period around the current stance of the economy, obtaining new (solution) coefficient matrices in each period and then calculating the values of the endogenous variables for the next period. In addition to being more accurate than standard linear solution approaches in most cases when doing simulations, the additional advantage of the DL is that it can be used to come up with a non-linear Kalman filter. Indeed, likelihood estimation on non-linear models have been a challenge, yet their policy-relevance only increases, especially after the Great Recession. Impulse response analysis and applying our non-linear Kalman to the US data do show how different estimates can be improved when the economy is far away from the steady state.

Despite the promising results, future work is needed for fine-tuning the new solution approaches. In addition, seeing how strongly their advantages spill over to bigger and richer dynamic macro-financial models, would be important for seeing implications about real-world questions of policy-makers and researchers.

References

[1] An, S. and Schorfheide, F., 2007. Bayesian analysis of DSGE models. *Econometric reviews, 26*(2-4), pp.113-172.

[2] Armstrong, J., Black, R., Laxton, D. and Rose, D., 1998. A robust method for simulating forward-looking models. *Journal of Economic Dynamics and Control, 22*(4), pp.489-501.

[3] Benes, J., Clinton, K., George, A., Gupta, P., John, J., Kamenik, O., Laxton, D., Mitra, P., Nadhanael, G.V., Portillo, R. and Wang, H., 2017. *Quarterly Projection Model for India; Key Elements and Properties* (No. 17/33). International Monetary Fund.

[4] Blanchard, O. and Kahn, C., 1980. The Solution of Linear Difference Models under Rational Expectations. *Econometrica, 48*(5), pp.1305-1311.

[5] Christiano, L.J., 2002. Solving dynamic equilibrium models by a method of undetermined coefficients. *Computational Economics, 20*(1-2), pp.21-55.

[6] Claessens, S. and Kose, M.A., 2018. *Frontiers of macrofinancial linkages.* Bank for International Settlements. No.95

[7] Driscoll, J.C. and Holden, S., 2014. Behavioral economics and macroeconomic models. *Journal of Macroeconomics, 41*, pp.133-147.

[8] Evans, R. and Phillips, K., 2015. *Linearization about the Current State: A Computational Method for Approximating Nonlinear Policy Functions during Simulation* (No. 2015-02). Brigham Young University, Department of Economics, BYU Macroeconomics and Computational Laboratory.

[9] Fair, R.C. and Taylor, J.B., 1983. Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models. *Econometrica*, pp.1169-1185.

[10] Fernndez-Villaverde, J., Rubio-Ramrez, J.F. and Schorfheide, F., 2016. Solution and estimation methods for DSGE models. In *Handbook of Macroeconomics* (Vol. 2, pp. 527-724). Elsevier.

[11] Gali, J., 2007. Monetary policy, inflation, and the business cycle. *Princeton UP*.

[12] Ize, A. and Yeyati, E.L., 2006. Financial de-dollarization: is it for real?. In *Financial Dollarization* (pp. 38-63). Palgrave Macmillan, London.

[13] Laxton, D., Rose, D. and Tambakis, D., 1999. The US Phillips curve: The case for asymmetry. *Journal of Economic dynamics and Control, 23*(9-10), pp.1459-1485.

[14] Smets, F. and Wouters, R., 2007. Shocks and frictions in US business cycles: A Bayesian DSGE approach. *American economic review, 97*(3), pp.586-606.

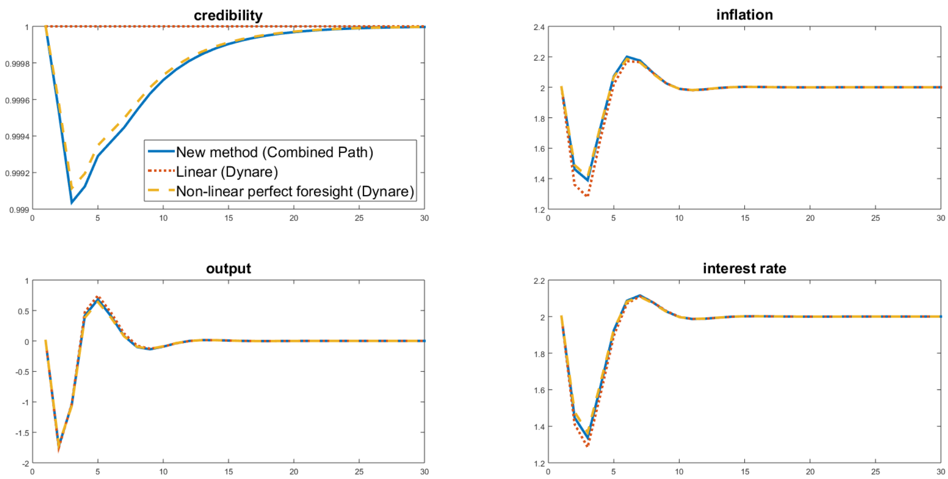
[15] Schorfheide, F., 2016, Particle Filters. *University of Pennsylvania.*

[16] Tvalodze, S., Mkhatrishvili, S., Mdivnishvili, T., Tutberidze, D. and Zedginidze, Z., 2016. National Bank of Georgia. *The National Bank of Georgia's Forecasting and Policy Analysis System.*

A Appendix - IRF figures

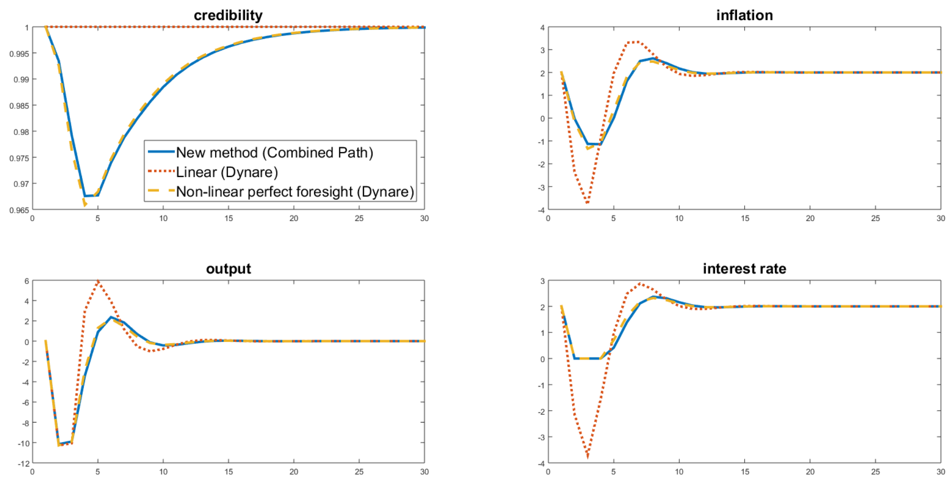
A.1 Figures for Combined Path

Figure 7: Impulse responses for 1% negative output shock (in the first period)



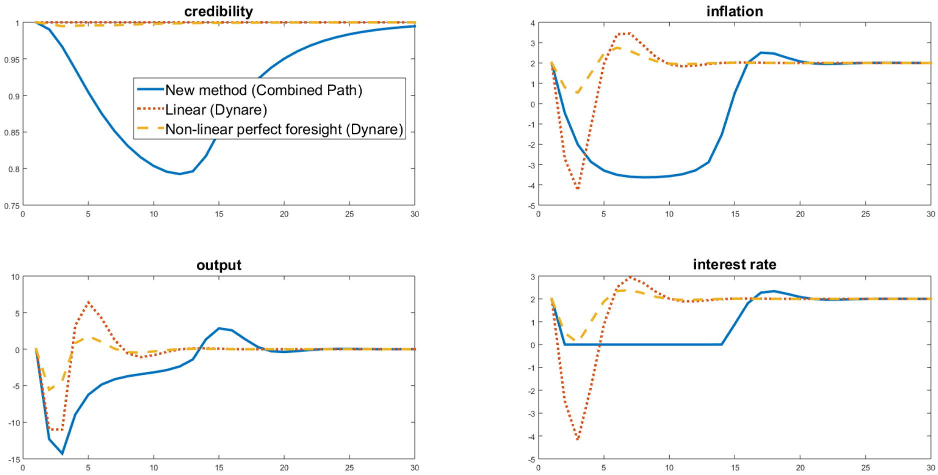
Source: Authors calculations

Figure 8: Impulse responses for a -10% output gap scenario (for the first two periods)



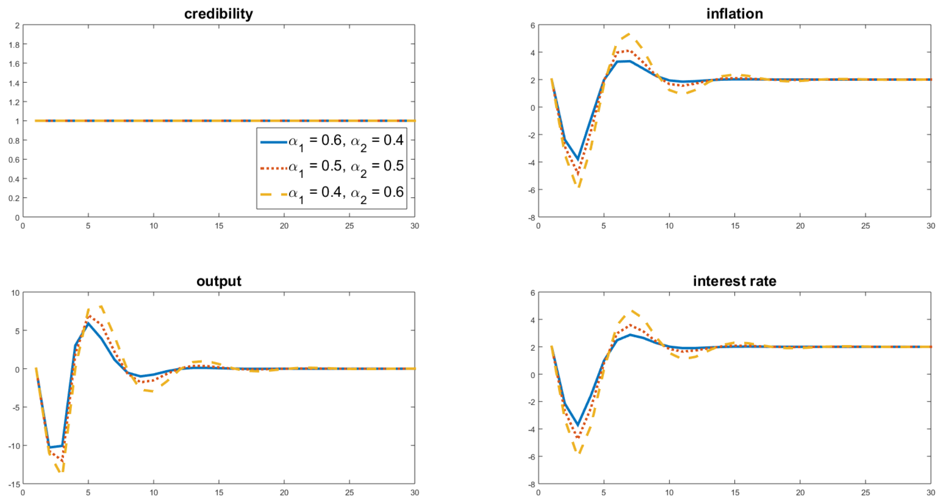
Source: Authors calculations

Figure 9: Impulse responses for a -10% output gap scenario (for the first two periods) augmented with an additional percentage point shock to the output gap



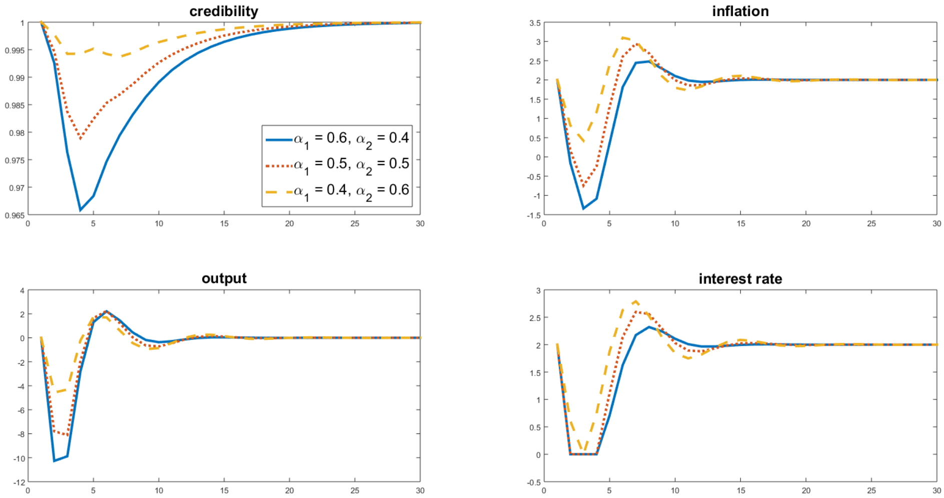
Source: Authors calculations

Figure 10: Impulse responses for a -10% output gap scenario (for the first two periods) across different degrees of habit formation - linear approach



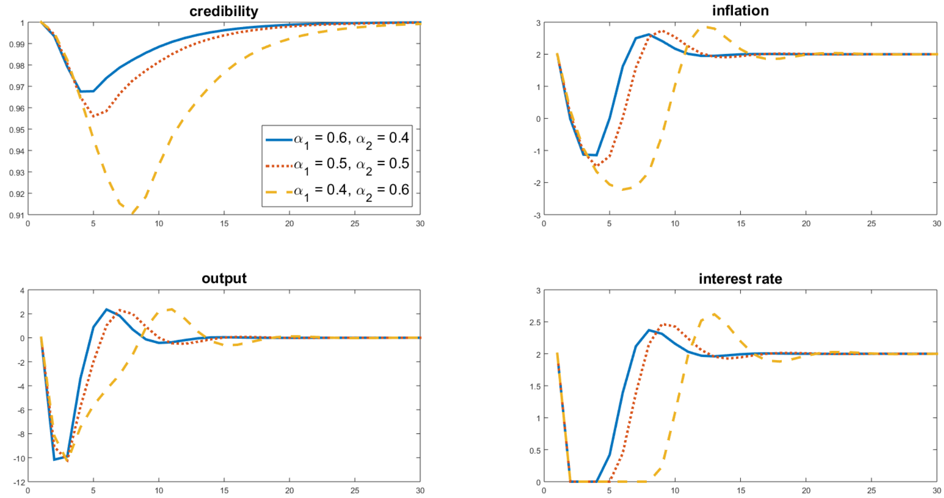
Source: Authors calculations

Figure 11: Impulse responses for a -10% output gap scenario (for the first two periods) across different degrees habit formation - EP approach



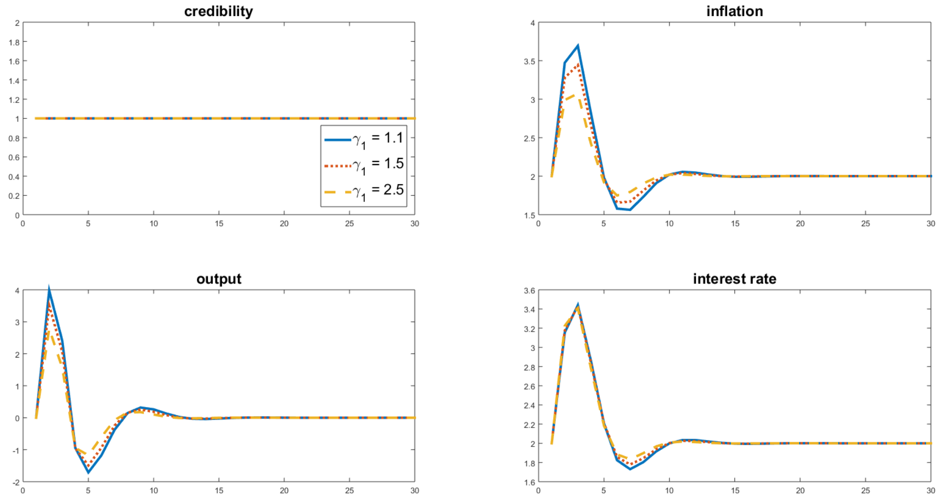
Source: Authors calculations

Figure 12: Impulse responses for a -10% output gap scenario (for the first two periods) across different degrees of habit formation - CP approach



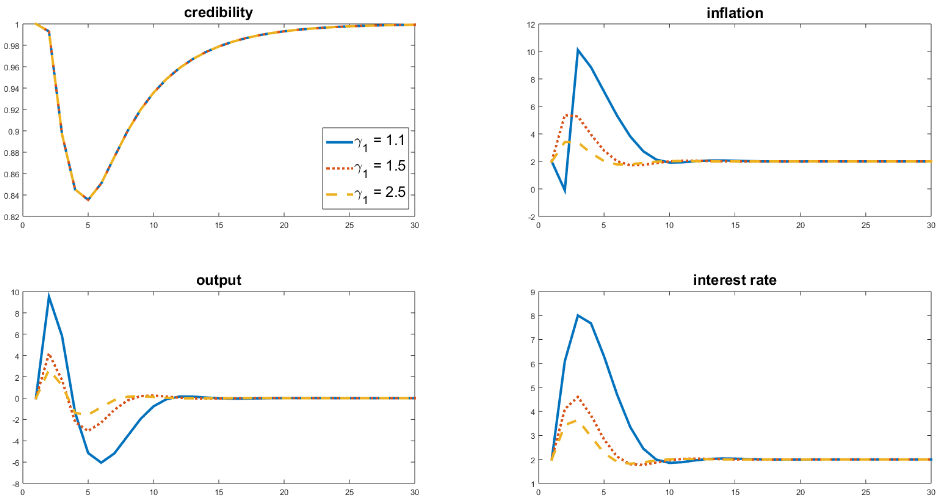
Source: Authors calculations

Figure 13: Impulse responses for positive 2% output shock (in each of the first two periods) across different degrees of policy reaction to inflation - linear approach



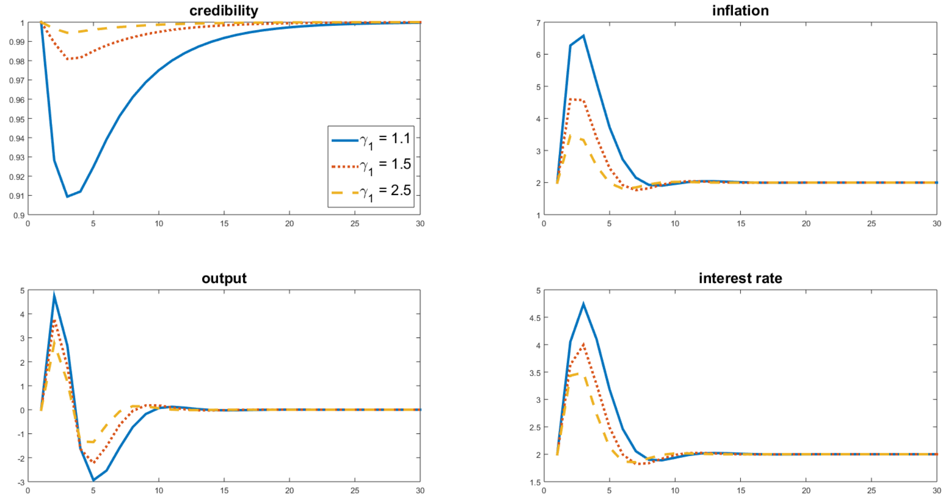
Source: Authors calculations

Figure 14: Impulse responses for positive 2% output shock (in each of the first two periods) across different degrees of policy reaction to inflation - EP approach



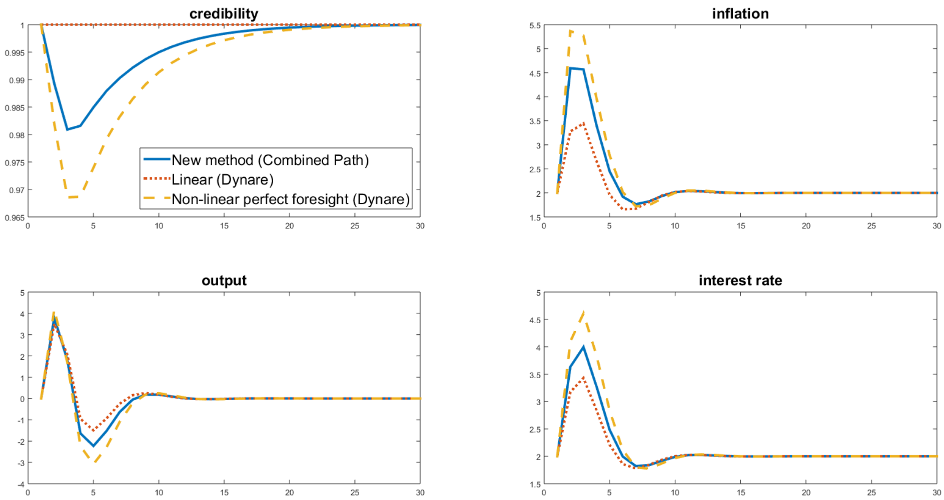
Source: Authors calculations

Figure 15: Impulse responses for positive 2% output shock (in each of the first two periods) across different degrees of policy reaction to inflation - CP approach



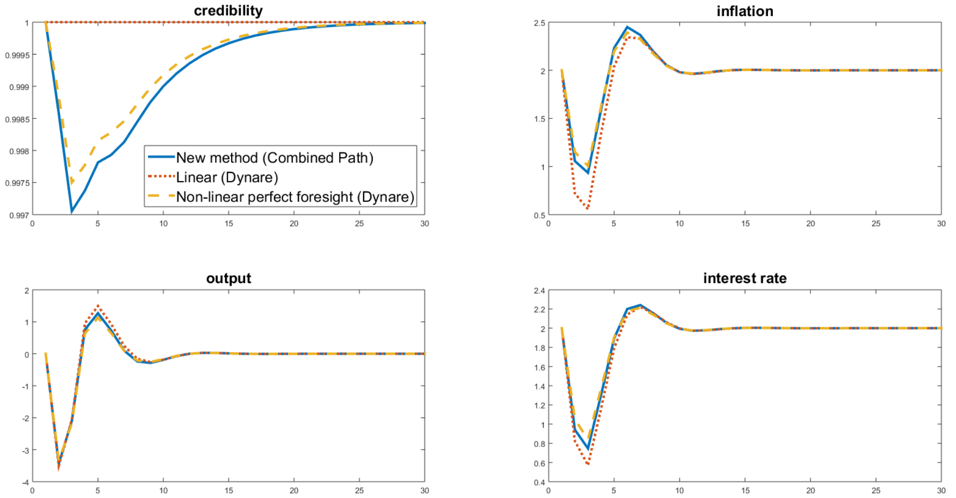
Source: Authors calculations

Figure 16: Impulse responses for positive 2% output shock (in each of the first two periods)



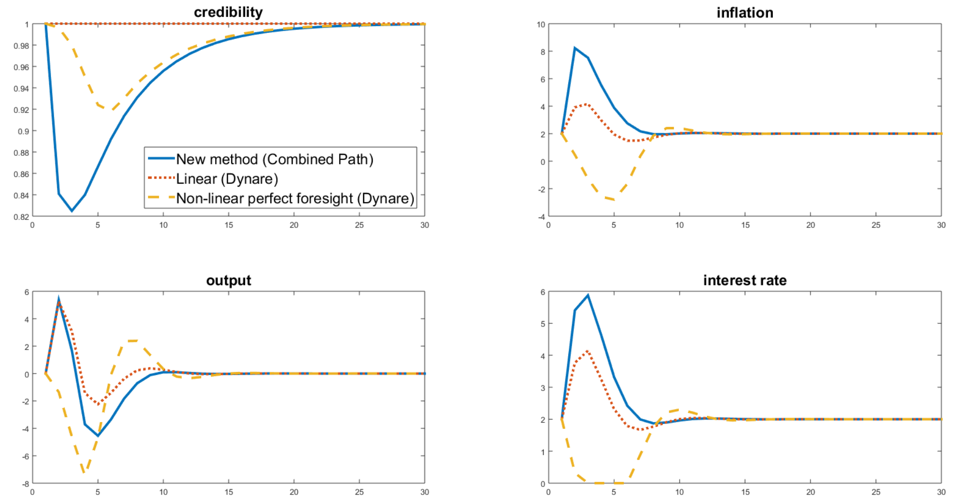
Source: Authors calculations

Figure 17: Impulse responses for negative 2% output shock (in each of the first two periods)



Source: Authors calculations

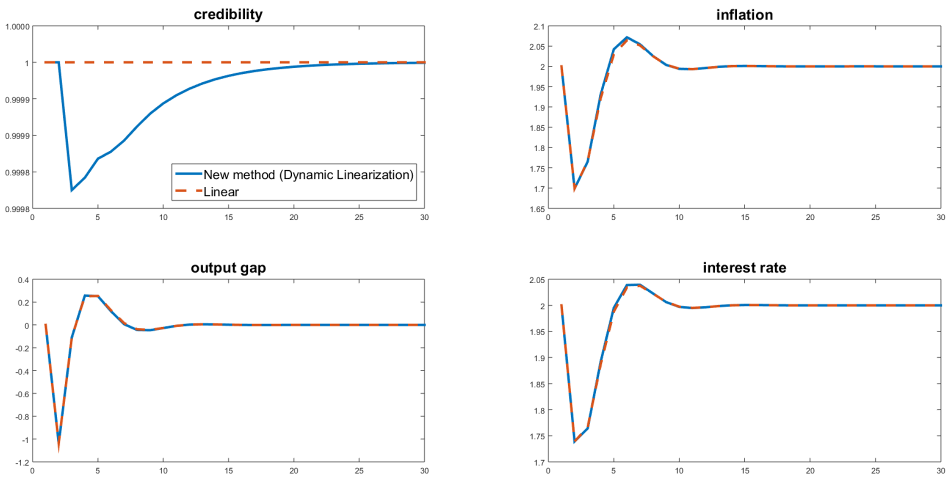
Figure 18: Impulse responses for positive 3% output shock (in each of the first two periods)



Source: Authors calculations

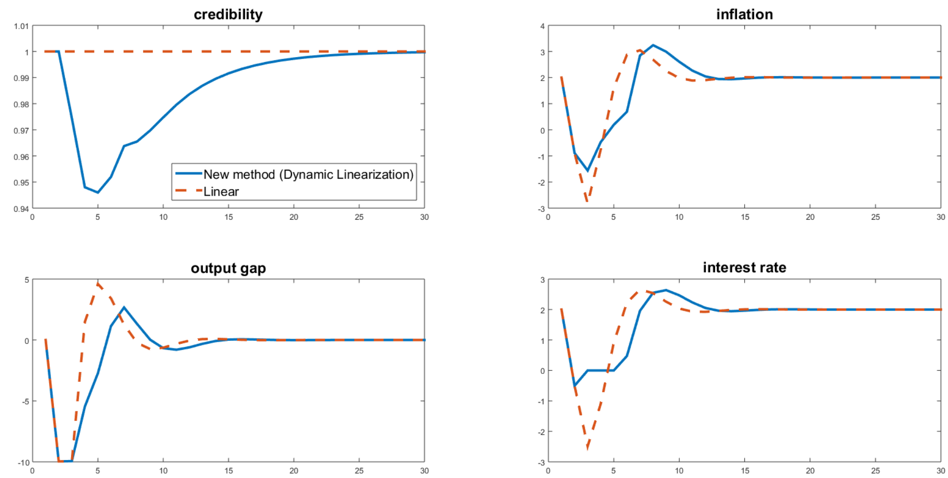
A.2 Figures for Dynamic Linearization

Figure 19: Impulse responses for 1% negative output shock in the first period



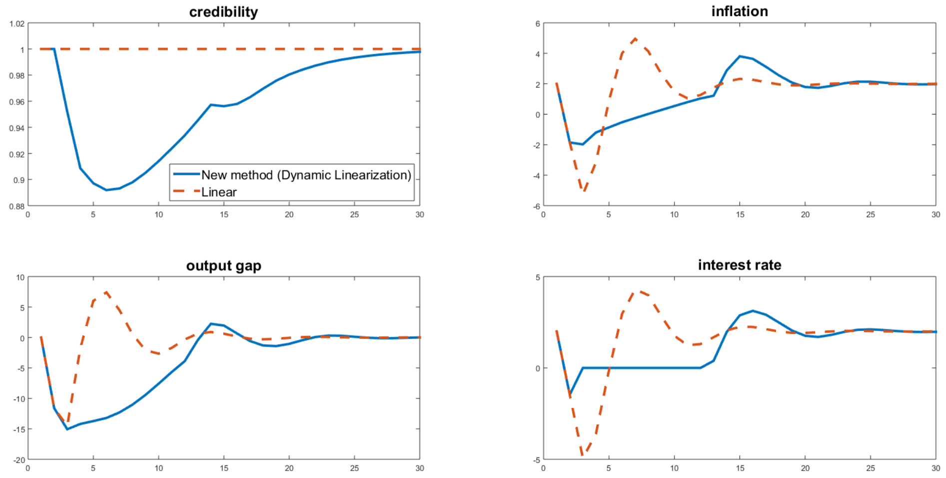
Source: Authors calculation

Figure 20: Impulse responses for a -10% output gap scenario (for the first two periods)



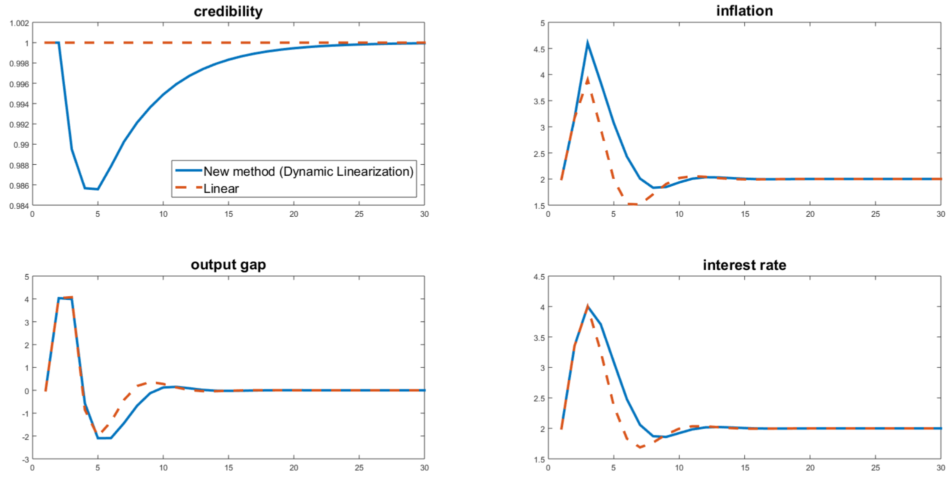
Source: Authors calculation

Figure 21: Impulse responses for a -10% output gap scenario with backward-looking IS curve



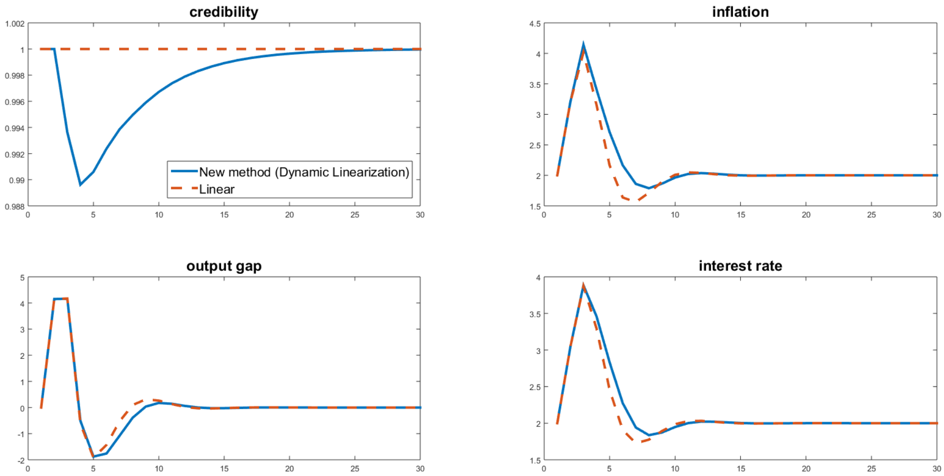
Source: Authors calculation

Figure 22: Impulse responses for anticipated 2% output shock (in each of the first two periods)



Source: Authors calculation

Figure 23: Impulse responses for unanticipated 2% output shock (in each of the first two periods)



Source: Authors calculation

Distributed by the National Bank of Georgia.

Available at www.nbg.gov.ge

National Bank of Georgia
Macroeconomic Research Division
Sanapiro st. #2, 0114, Tbilisi, Georgia
Phone: +995 32 2406531
www.nbg.gov.ge
Email: research@nbg.gov.ge

საქართველოს ეროვნული ბანკი
National Bank of Georgia