Model-Based Ex Post Evaluation of Monetary Policy*

Eyal Argov, Alon Binyamini, Eliezer Borenstein, and Irit Rozenshtrom
Research Department, Bank of Israel

We present a model-based methodology to conduct an ex post evaluation of monetary policy decisions, by testing whether alternative policy decisions could have brought a Pareto improvement in terms of inflation and output volatilities. This involves simulations of counterfactual scenarios under alternative monetary policy shocks, and computation—for each such simulation—of the root mean square (RMS) of the inflation and output gaps during and following the evaluated year. It is then possible to compare the actual RMS with simulation-based frontiers, with each frontier reflecting different constraints on interest rate volatility, which can be viewed as a third objective variable. The actual RMS is also compared with the counterfactual RMS that would have been obtained under the case of no policy shocks. Such comparisons enable testing whether monetary policy shocks were "ex post efficient." The methodology is implemented in an evaluation of Bank of Israel policy decisions during the years 2001–11. The implementation shows several distinct sets of years: years in which actual RMSs were close to the efficient frontiers and years in which they were distant from them; years in which monetary policy shocks led to an absolute improvement in economic outcomes (by reducing the RMSs of both inflation and output gaps) or an absolute deterioration; and years in which policymakers faced a trade-off between all three objective variables or

*The authors would like to thank Guy Segal, Roni Frish, participants at the Bank of Israel Research Department seminar, two anonymous referees, and the co-editor for their valuable comments. Author e-mails: eyal.argov@boi.org.il, alon.binyamini@boi.org.il, eliezer.borenstein@boi.org.il, and irit.rozenshtrom@boi.org.il.

between a subset of the variables. For most evaluated years, the results seem qualitatively robust, considering the uncertainty in the historical shocks extracted, as well as in alternative definitions for the output gap.

JEL Codes: C54, E37, E52, E58.

1. Introduction

The purpose of this paper is to offer an analytical framework for an ex post evaluation of the efficiency of monetary policy decisions. A motivation for such a framework is discussed by Svensson (2012), who argues that accountability and regular evaluation of the central bank's decisions are important incentives in motivating independent and powerful central banks. To that end, he encourages the development and application of reasonable principles and appropriate methods.

To the best of our knowledge, there has been little work done on principles and methods for such an ex post evaluation of monetary policy decisions. The vast literature on optimal policy *rules* focuses on the general functioning of decision makers, i.e., evaluating the policy rule, rather than the actual decisions.[1] One exception is Svensson (2012), who proposes a framework for evaluating a specific policy decision. This involves calculating the expected root mean squares (RMSs) of the inflation and output gaps ex ante, as they were projected when the interest rate decisions were made. Expected RMSs are also calculated for alternative future interest rate paths using a DSGE model and then compared with the baseline case. Such a procedure allows the quantifying of the trade-off between stabilizing inflation and output, and may show at times that the expected RMSs of both inflation and output could have been reduced, and therefore that the policy decisions were ex ante inefficient.

The approach taken here differs from that of Svensson (2012) in two main ways: First, the approach can be applied in the evaluation of a sequence of policy decisions, say over a period of a year, and not just of a projected path associated with a single decision.

[1] Using the terminology of DSGE models, the latter includes not only the systematic part of the rule but also policy "shocks," i.e., non-systematic deviations from the rule.

Second, and more importantly, we employ an approach that evaluates past monetary policy decisions based on ex post information, which was not always available in real time. We nevertheless believe that it is of interest to examine what would have been the counterfactual outcomes under alternative decisions, what would have been the best economic results that policymakers could have delivered, and what are the measures that should have been taken to achieve those results. Moreover, a single model clearly cannot forecast certain developments that may be predictable using other tools. Hence, the model employed may regard such developments as unexpected shocks (revealed only ex post), whereas in reality they could have been predicted and taken into account by decision makers. Put differently, to some extent, ex post evaluation may point to conclusions relevant ex ante as well.

The proposed evaluation framework compares the actual volatility of two main objective variables—the inflation and output gaps—to their counterfactual volatility associated with alternative interest rate paths. The analysis uses a DSGE model, and the volatility is measured by the RMS of a variable during the evaluated period and its expected path for subsequent periods. It is the effect of policy decisions on future outcomes that motivates the inclusion of the expected path in the calculation of the RMS. The counterfactual interest rate paths are obtained by repeatedly hitting the model economy with randomly selected monetary policy shocks. One particular benchmark alternative path is the case of zero policy shocks, i.e., strictly following the interest rate rule. In addition, we numerically search for the efficient frontier, i.e., the combinations for which the RMSs of both objective variables cannot be simultaneously improved. We regard the "active component" of the policy (that is, the non-systematic policy shock) as "efficient" when actual RMSs are closer to the frontier, compared with those consistent with the zero-policy-shock path. In addition, the evaluation framework can be used to assess whether the actual policy was more in favor of stabilizing inflation or stabilizing activity.

The main underlying assumption of this evaluation framework is that, over short horizons, an active policy means choosing deviations from an interest rate rule, i.e., manipulating the monetary policy shocks, rather than changing the policy rule per se. This assumption stems from our view that the systematic component of monetary policy rules cannot, and should not, be changed frequently. One reason is that central banks do not normally publish their policy rule—certainly not on a quarterly basis. If anything, central banks publish an expected policy *path.* Therefore, considering counterfactual outcomes for alternative policy *rules* may not be particularly relevant. Since frequently changing the policy rule is not practical, and since the public presumably employs some form of assumed rule to form expectations with regard to future economic developments, we interpret active monetary policy to be the choice of short-run deviations from the rule, i.e., monetary policy "shocks." The driving forces behind such shocks may be (temporary) changes in the loss function's weights (such as a change in the relative importance assigned to stabilizing inflation or output, as well as the weight assigned to interest rate smoothness) or considerations that are not systematically incorporated into the rule (for example, precautionary reaction to risk). It should be noted that there are, of course, alternative interpretations for the deviations from an interest rate rule, which are not consistent with our methodology; these include real-time measurement errors in the variables entering the rule and errors resulting from misspecification of the "policy rule."

Our approach thus focuses on the discretionary, or the non-systematic, part of the monetary policy. Such an approach is consistent with the view expressed by Bernanke and Mishkin (1992), who stress that rules do not always allow for responding to unforeseen circumstances. It is also consistent with Svensson (2005), who argues that monetary policy which incorporates judgment may perform better than simply following an instrument rule. It is exactly this view of the non-systematic, or discretionary, component that makes our evaluation framework relevant.

In contrast to our focus on the non-systematic component, Taylor (2011a, 2011b) emphasizes the important stabilizing role of rule-based policy, that is, of the systematic part of the policy reaction function. This view stresses the expectations channels of the monetary policy transmission mechanism.[2] In this context, it is

[2]Taylor (2011b) argues that the first decade of the 2000s was characterized by great deviations from the rule-based policy, a rule that was successful in stabilizing economic activity, thus attaining the so-called Great Moderation. He further argues that it was this great deviation which was responsible for the Great Recession that followed. Note that this view fits well into our approach.

worth noting that we assume that the policy shocks, which reflect the discretionary component of the policy, do not impair the credibility of the policymaker. Thus, the rule-based part of the policy continues to have an effective stabilizing role.

In the proposed framework, the criterion for policy evaluation focuses on inflation- and output-gap volatilities. However, it is common to suppose that monetary policy attributes some value to interest rate smoothness as well. Therefore, we compute different efficient frontiers, corresponding to different limitations on interest rate volatility. This allows the analysis to introduce, in some sense, a third argument into the implicit loss function.

To demonstrate an implementation of the framework, we employ the Bank of Israel's medium-scale DSGE model (MOISE)[3] to evaluate the conduct of monetary policy in Israel during the years 2001–11. During this decade, inflation in Israel fluctuated around the center of the target range, though there were sizable deviations above and below the midpoint. For most of the period, the (model-based) output gap was negative (2 percent on average), which in retrospect can be viewed as an indication that there was considerable room for further monetary easing, namely policy shocks, that may have improved economic outcomes. The implementation reveals several distinct sets of years: years in which actual RMSs were close to the efficient frontier and years in which they were distant; years in which monetary policy shocks led to an absolute improvement in economic outcomes or to an absolute deterioration; and years characterized by aggressive policy shocks, which were usually aimed at narrowing the output gap at the expense of more volatile inflation. The implementation is followed by robustness checks in which we consider the sensitivity of the conclusions to the extraction of the historical shocks and to alternative definitions for the output gap. For most evaluated years, these robustness checks suggest that the results remain qualitatively unchanged and, accordingly, the conclusions remain valid.

The rest of the paper is organized as follows. Section 2 formally describes the framework for policy evaluation. Section 3 demonstrates an implementation of the framework, by evaluating monetary

[3]See Argov et al. (2012).

policy in Israel during the decade from 2001 to 2011. In that section we will only elaborate on a few distinct years. In section 4 we present two robustness analyses—with respect to alternative extractions of historical shocks and with respect to alternative definitions of the output gap. Finally, section 5 offers some concluding remarks—both on the framework and on the evaluation of the Bank of Israel's monetary policy during the first decade of the millennium as a whole.

2. The Evaluation Framework

2.1 A Model

This section describes the formal framework. In order to apply the model-based evaluation framework, one has to choose a model to be employed both for the extraction of historical shocks that are consistent with the observed history and for the simulation of counterfactual scenarios. Monetary policy must be explicitly specified in this model, using a rule that includes an exogenous shock. These restrictions are very general so that any commonly used DSGE model is adequate.

For the demonstration of the framework in section 3 below, we use the Bank of Israel's medium-scale DSGE model, MOISE, described in Argov et al. (2012). The model follows along the lines of the European Central Bank's New Area-Wide Model and the Riksbank's Ramses model (see Christoffel, Coenen, and Warne 2008 and Adolfson et al. 2007, respectively).[4] The economic entities in the model are households, a production sector (producers and importers of intermediate goods, producers of final goods, and exporters), a fiscal authority, and an inflation-targeting central bank whose policy tool is the nominal interest rate ($\widehat{r}_t$). The interest rate rule in MOISE takes the following form (a hat over a variable denotes logarithmic deviation from a steady-state value or from a trend):

$$\begin{aligned}\widehat{r}_t = (1-\phi_R)\left[\widehat{rr}_t^{fwd} + \widehat{\overline{\pi}}_t + \phi_\Pi\left(\widehat{\pi}_t^{CB} - \widehat{\overline{\pi}}_t\right) + \phi_y \widehat{y}_t^{GAP} + \phi_{\Delta S}\Delta S_t\right] \\ + \phi_R \widehat{r}_{t-1} + \eta_t^R, \qquad (1)\end{aligned}$$

[4]Other similar central bank models include Brubakk et al. (2006) for Norway, Murchison and Rennison (2006) for Canada, Beneš et al. (2009) for New Zealand, and Seneca (2010) for Iceland.

where the inflation to which the rule responds is

$$\widehat{\pi}_t^{CB} = E_t \left[\widehat{\pi}_{C,t-2} + \widehat{\pi}_{C,t-1} + \widehat{\pi}_{C,t} + \widehat{\pi}_{C,t+1} \right], \tag{2}$$

with $\widehat{\pi}_{C,t}$ being the quarterly inflation of the consumer price index (CPI) in period t.

Thus, the interest rate is driven by deviations of (expected) inflation from the inflation target $\left(\widehat{\pi}_t^{CB} - \widehat{\overline{\pi}}_t\right)$, the output gap $\left(\widehat{y}_t^{GAP}\right)$—defined below as the deviation of output from a technology-driven trend—and nominal depreciation (ΔS_t). The variable $\widehat{rr}_t^{fwd}$ is the long-run real interest rate, and $\widehat{\overline{\pi}}_t$ is the time-varying inflation target. The equation includes a policy shock, η_t^R, which is assumed to follow a white-noise process.

In this paper we define the output gap as the logarithmic deviation of the observed output (Y_t) from a technology-driven trend:

$$\widehat{y}_t^{GAP} \equiv \log \frac{Y_t}{Z_t \varepsilon_t} - \log y, \tag{3}$$

where Z_t and ε_t are, respectively, permanent and transitory technology shocks (external shifts in the total factor productivity) and y is the steady state of the productivity-adjusted output. This definition of the output gap, which will also be used as an objective variable in the policy evaluation, is a measure of resource utilization. In MOISE it essentially accounts for deviations of production inputs (capital and labor) from some (unobserved) trend. Hence, our output-gap measure may be interpreted as reflecting demand pressures and rigidities such as “time to build,” rather than an alternative measure that uses only the permanent component of technology as the trend. There are some alternative measures for the output gap that are used in the New Keynesian literature, particularly in the literature on optimal monetary policy. These include the gap from the hypothetical flexible-price output or from perfectly competitive output. Apart from the complexities involved in calculating these theoretical measures, we suspect that using them in practice for policy evaluation might make the results more dependent on model assumptions. Svensson (2012), who uses output deviations from HP trend, stresses the importance of having several points of view with respect to resource utilization. In line with this view, subsection 4.2 presents robustness analysis with respect to alternative measures for the output gap.

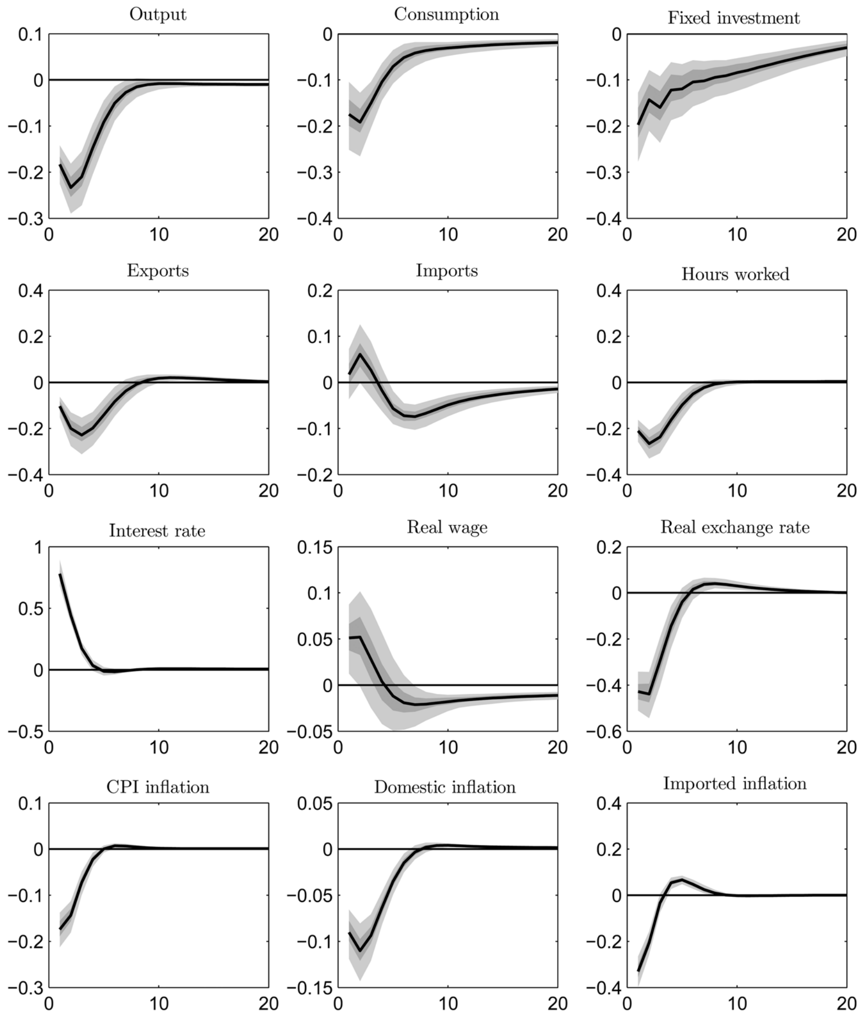
2.2 The Monetary Policy Shock

Since the non-systematic part of the policy rule, namely the monetary policy shock, is at the center of our evaluation approach, in this subsection we briefly discuss the effect of a monetary policy shock in the model. This shock will be manipulated within our evaluation framework, in order to simulate alternative, counterfactual outcomes. Figure 1 presents the impulse response function (IRF) for several key variables following a monetary policy shock, η^R.

As can be seen from figure 1, an innovation to the interest rate rule (1) of one standard deviation triggers an immediate increase of the interest rate of 0.75 percentage points. This increase causes output to fall by approximately 0.2 percent. Note that output reaches its lowest point only after two quarters and gradually converges back to its trend within two years. CPI inflation falls on impact by 0.2 percentage points, with an accumulated effect of about 0.4 percentage points, one year after the shock.

It is interesting to compare the model's IRF to those reported for similar models of other economies, such as in Christoffel, Coenen, and Warne (2008) and Adolfson et al. (2007) for the euro area, Adolfson et al. (2008) for Sweden, and Beneš et al. (2009) for New Zealand. This comparison leads to four general observations: (i) the size of the shock in MOISE is typically larger by a scale of 1.5 to 3 (reflecting a larger estimated standard deviation of the interest rate shock); (ii) while the effect on output in MOISE is typically smaller, mainly due to the lower sensitivity of investment, the effect on inflation is larger due to a faster exchange rate pass-through (along with higher import intensity); (iii) the output reaction in MOISE is faster and less hump shaped, with the strongest effect only two quarters following the shock, as compared with three or four quarters in other economies; and finally, (iv) the effect of the shock has a shorter duration in MOISE (two years as compared with five years in other economies).

Figure 1. Impulse Response to an Interest Rate Shock



Notes: Shock of one standard deviation. Solid line: mean of impulse response. Gray area: 70 and 90 percent highest interval of impulse response. Real variables: percentage deviation from steady state. Inflation: percentage-point deviation from steady state. Interest rate: annualized percentage-point deviation from steady state.

2.3 Formal Description of the Algorithm

This section lays out a step-by-step description of the evaluation framework. The outline of the algorithm is followed by a discussion of the typical output of the evaluation exercise.

Step 1: Estimate the Historical Shocks. The Kalman filter is used to extract historical shocks η_t for all periods t in the sample, given a set of observed variables.[5] η_t is a vector of all the model's i.i.d. shocks including the monetary policy shock, η_t^R. Note that since the historical shocks are not uniquely identified (there are infinitely many combinations of shocks consistent with the observable variables), we employ the Kalman filter to smooth out the conditional mean of the distribution of shocks' estimates. In section 4.1 we present a robustness test, accounting for the uncertainty associated with the actual realization of these historical shocks.

Step 2: Choose the Period of Evaluation and Calculate the Actual RMSs. The period of evaluation does not necessarily have to be a year. Nonetheless, for our implementation below we choose to work with calendar years, for several reasons: (i) it is a natural reference period; (ii) it is short enough so that monetary policy may be thought of as choosing deviations from some rule and not as choosing different rules; and (iii) it is long enough for policy to have an economically significant effect on the target variables, as is evident from the impulse response presented in figure 1.[6]

For the chosen period, calculate the RMSs of the variables assumed to be targeted by monetary policy. In our implementation, we assume the policy criterion to be a loss function mainly based on the RMS of the deviations of the inflation rate from its target (inflation gap, $\widehat{\pi}_t^{GAP} = \widehat{\pi}_{C,t} - \widehat{\overline{\pi}}_t$) and deviations of output from a trend consistent with technological capacity (i.e., the output gap, $\widehat{y}_t^{GAP}$). Formally, we let t be the first quarter of the evaluated year, and calculate

[5] Throughout the paper, we actually use the smoothing process—that is, estimating unobservable shocks based on the entire sample.

[6] Israel has a relatively rapid transmission mechanism. Therefore, in applying this framework to other economies, one might choose to evaluate periods of two to three years.

$$RMS(\widehat{\pi}_{t,t+7}^{GAP}) = \sqrt{\frac{1}{8}\left[\begin{array}{c} (\widehat{\pi}_{C,t}-\overline{\pi}_t)^2+\cdots+(\widehat{\pi}_{C,t+3}-\overline{\pi}_{t+3})^2 \\ +E\,(\widehat{\pi}_{C,t+4}-\overline{\pi}_{t+4})^2+\cdots+E\,(\widehat{\pi}_{C,t+7}-\overline{\pi}_{t+7})^2 \end{array}\right]} \quad (4)$$

and

$$RMS(\widehat{y}_{t,t+7}^{GAP}) = \sqrt{\frac{1}{8}\left[\left(\widehat{y}_t^{GAP}\right)^2+\cdots+\left(\widehat{y}_{t+3}^{GAP}\right)^2+\left(E\widehat{y}_{t+4}^{GAP}\right)^2+\cdots+\left(E\widehat{y}_{t+7}^{GAP}\right)^2\right]}. \quad (5)$$

In equations (4)–(5), E is the mathematical expectation operator, conditional on the information available at the last quarter of the evaluated year, $t+3$, so that the RMS includes a forecast for the respective variable in the following year. One reason for taking the expectations rather than future realizations (ex post) in calculating the "actual" RMS is to avoid including the effects of future policy decisions in the analysis, thus focusing on policy decisions in the period under evaluation only. Also, the further ahead into the future the realization of shocks is, the less we would expect policymakers to anticipate them and act accordingly. Finally, this way the evaluation of policy for a historical period is similar to an evaluation one would have performed in the end of that period, bringing us closer to an evaluation performed in real time.[7]

At this stage we also calculate the within-year standard deviation of the interest rate:

$$\sigma_{t-1,t+3}^r = \sqrt{\frac{1}{5}\left[(r_{t-1}-\overline{r})^2+(r_t-\overline{r})^2+\cdots+(r_{t+3}-\overline{r})^2\right]}, \quad (6)$$

where r_t denotes the annualized nominal interest rate and $\overline{r}$ is the within-year average.[8]

[7]Two reasons why the results might be different, compared with those based on real-time evaluation, are data revisions and the Kalman smoother which exploits information inherent in the entire sample, available only to the ex post researcher. We remark on the issue of data vintages further below.

[8]In fact, starting in $t-1$, expression (6) also includes the last quarter of the preceding year, capturing the idea that the evaluated year is part of a

Step 3: Run a Counterfactual Simulation with Zero Monetary Policy Shocks. Starting from the smoothed state in the last quarter before the evaluated period, a counterfactual scenario is simulated based on the vectors of extracted shocks from step 1, except for the monetary policy shocks which are set to zero in every quarter of the evaluated period. All shocks for the periods following the evaluated one are set to zero as well. The result suggests what would have been the case had monetary policymakers not deviated from the estimated interest rate rule.

Using the simulation results, the RMSs of the main objective variables are calculated, as in equations (4) and (5). Comparing the gaps between the resulting counterfactual RMSs with the actual ones, from step 2, reflects the contribution of the policy's nonsystematic component.

Step 4: Run Numerous Counterfactual Simulations with Randomly Selected Monetary Policy Shocks, and Estimate Policy Frontiers. Run numerous simulations based on counterfactual, alternative monetary policy shocks. In our implementation, 5,000 four-period vectors of monetary policy shocks are drawn from the estimated distribution of the monetary policy shock, $N\left(0, \sigma^2_{\eta^R}\right)$. For each draw, a counterfactual simulation is run with the vector of drawn monetary policy shocks and the remaining smoothed shocks extracted by step 1. For each simulation, the RMSs are calculated for the objective variables, as in equations (4) and (5).

In addition, we estimate the feasible policy frontier—the combinations of RMSs for which it is impossible to improve the RMSs of both the output gap and the inflation gap. Moreover, in order to treat interest rate volatility as an additional consideration of policymakers, we distinguish between different frontiers, each corresponding to a different ceiling on the within-year interest rate standard deviation, as calculated in equation (6). Note that the frontiers are estimated using a search algorithm (presented in the appendix), such that each frontier is based on additional simulations

continuum. That is, when deciding on the interest rate for the first quarter of the calendar year, the central bank is also constrained by the interest rate of the previous quarter. For the same reason, the "within-year average" also includes that last quarter of the preceding year.

in addition to the 5,000 random draws from the estimated distribution.[9]

Step 5: Generate a Diagrammatic Representation of the Results. In our implementation below, we present the evaluation results using two types of diagrams: (i) a scatter diagram that presents the counterfactual RMSs of the objective variables, along with feasible frontiers; and (ii) the interest rate paths of the simulations associated with one of the frontiers.

The first type of diagram (see, for example, panel A in figure 4) presents the RMSs of the inflation gap (horizontal axis) and the output gap (vertical axis). The large point labeled "Actual" represents the actual RMSs (computed in step 2); the large point labeled "Zero Shocks" is the RMS from the zero-shock counterfactual simulation (step 3); and the small points are the RMSs from the random simulations (step 4). The diagram also presents three feasible frontiers (marked by F) of the gaps' RMSs—limited by once $F(1\mathrm{X}\sigma^r)$, twice $F(2\mathrm{X}\sigma^r)$, and three times $F(3\mathrm{X}\sigma^r)$ the sample average of within-year interest rate standard deviation. In some sense, this introduces a third dimension to the diagram.

The diagram provides the following observations:

- The diagram indicates whether actual policy shocks led to outcomes that are closer to or farther away from the frontier, compared with the benchmark of zero policy shocks, and whether they reduced the RMSs of the inflation gap or the output gap.
- If the actual result is closer to the frontiers, compared with the counterfactual scenario of the zero-policy alternative, the diagram can indicate whether it reflects a liberal or a conservative attitude of the central bank.

[9]In both procedures (the counterfactual simulations and the frontier search), we discard draws leading to non-feasible negative interest rate paths. Nevertheless, we can report that for most evaluated years, this limitation is not binding (less than 0.5 percent of the randomly drawn draws are discarded). Only for the years 2009–10, during which the actual interest rate was approaching the zero lower bound, is this limitation binding, leading to a 5–25 percent discard rate. Since our (log-linear) model does not address the non-linearity associated with the zero lower bound, we take caution when interpreting the results for those years.

- The distinguished frontiers illustrate the quantitative implication of a lesser degree of interest rate smoothness.
- The diagram indicates whether the actual (smoothed) policy shocks delivered RMS results inside or outside the cluster cloud (generated by alternative policy shocks, randomly drawn around the counterfactual case of zero policy shock). In some sense, this indicates whether the actual policy actions were "significantly" different (for better or for worse) from the alternative of strictly following the monetary policy rule.

The second type of diagram (see, for example, panel B in figure 4) depicts the interest rate paths that generate the results along the frontier (the dashed lines). To economize on space, we only present the diagram related to the frontier limited by twice the sample interest volatility $F(2\text{X}\sigma^r)$. For comparison, the diagram also shows the actual path (the solid line) and the zero-policy-shock path (the line with the circle marker). This type of diagram reveals whether the actual path chosen in the evaluated period was above, below, or within the mass of efficient paths (associated with certain constraints on interest rate volatility).

For each evaluated year, we also report the within-year standard deviation of the interest rate, compared with its sample average (see, for example, the table below panel A in figure 4).

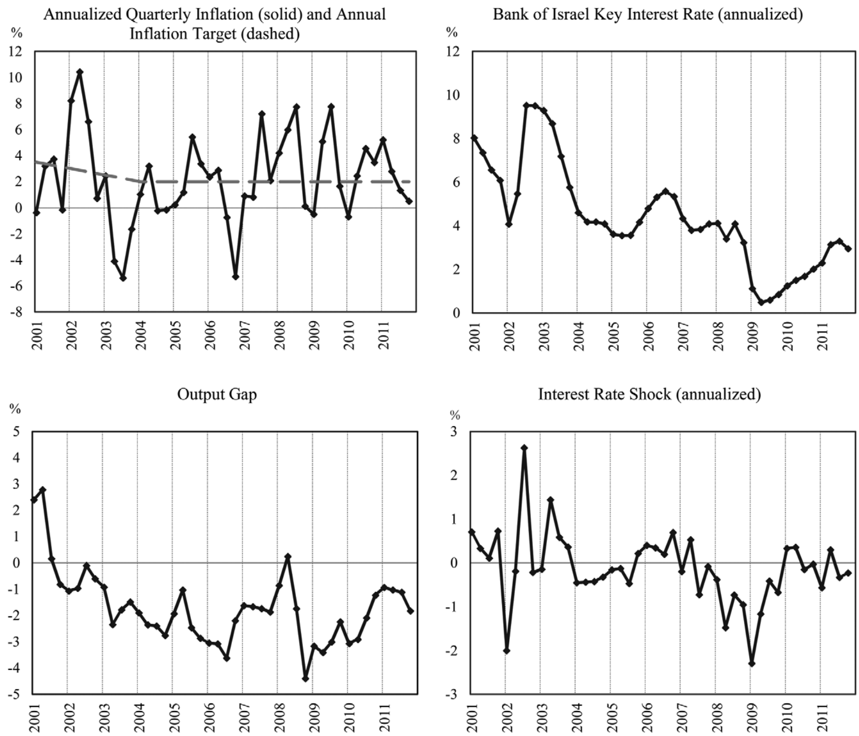
3. Implementation

This section presents an implementation of the evaluation framework to assess monetary policy in Israel during the first eleven years of the present millennium.

3.1 Historical Perspective

Before implementing the evaluation framework for specific years, it is worthwhile to briefly review the developments in the variables of interest during the relevant period, 2001–11. Figure 2 depicts the quarterly inflation rate, the output gap, the nominal interest rate, and the smoothed monetary policy shock consistent with equation (1).

Figure 2. The Main Variables of Interest, 2001–11



Inflation fluctuated around the midpoint of the target range, though with significant variance. Thus, the inflation rate was above target in the years 2002, 2005, 2007, 2008, 2009, 2010, and 2011 and below target in the years 2003, 2004, and 2006. The output gap was negative for most of the evaluated years, 2002–11. The recessionary downward trend of the gap during 2001–3 was due to a combination of three driving forces: a deteriorating security situation (the second “Intifada” which began in late 2000), the dot-com crash of 2000–1 (which had a large negative effect on the high-tech sector of the Israeli economy), and the impaired credibility of macroeconomic policy. The gap narrowed until mid-2008, when the Global Financial Crisis began, and started improving again in mid-2010.

The interest rate path is characterized by a downward trend, in part due to the disinflation process (see the top-left panel in figure 2) and in part due to the reduction in long-term real yields. However, for short horizons (up to one year) variance decomposition of the nominal interest rate, as presented in Argov et al. (2012), shows the important role of the monetary policy shock: 56 percent for one quarter and 17 percent for four quarters. Inspecting the smoothed monetary policy shock (bottom-right panel in figure 2) shows three outliers: (i) In early 2002, the interest rate was unexpectedly reduced by 2 percentage points (as part of an agreement with the government on expenditure cuts), while in the second part of that year, against the background of rising inflation, there was a sharp increase in the interest rate, once again through large policy shocks. (ii) The years 2005–6 were characterized by small but persistent positive policy shocks. This occurred in the aftermath of a large depreciation in 2005 which generated concern regarding a negative interest rate gap relative to the federal funds rate. The result was low inflation and output that year. (iii) The years 2008–9 were characterized by strong and persistent negative policy shocks, reflecting preemptive measures against the background of the Global Financial Crisis.

Before we turn to the evaluation of policy for some specific years, a note on the data used in the evaluation process is in order. The data used in this work are the updated data available at the time the analysis was conducted (April 2014). To the extent that some of the data (particularly National Accounts data) is prone to revisions, the issue of what data vintage one should use when evaluating past policy decisions may arise. Note also that aside from the issue of possible data revisions, the question arises also regarding the use of future data to infer about previous periods when applying the Kalman smoother. In our view, using the most up-to-date data seems appropriate given the ex post nature of the analysis, where one is interested in examining policy decisions in light of our best understanding of the actual, realized circumstances ex post. Note that if the focus is on the ex ante perspective for some past period, then it would be more appropriate to use the data that was available at that time—not only when applying the Kalman smoother but also for estimating the model. Of course, this would come at the cost of losing the information embedded in the data of the succeeding periods.

We implemented our evaluation framework for each of the years 2001–11. Subsections 3.2–3.4 elaborate on a few distinct years (presented in figures 3 to 5). In the concluding section we will provide some general observations from the full-decade evaluation.

3.2 Evaluation of 2003: Policy Shocks Increased the RMSs of Both Objective Variables

The year 2003 followed a period of stagflation that ended with a sharp rise in the interest rate to 9 percent in mid-2002. As a result, inflation declined in 2003 and fell well below the target during the second part of that year. In addition, slow growth led to a widening of the already negative output gap, which also contributed to a low inflation rate. After the realization that second-quarter inflation was too low, the interest rate was reduced to approximately 6 percent by the end of the year. According to the model, this reduction was smaller than required by the interest rate rule, implying that monetary policy shocks were in fact positive (see figure 2).

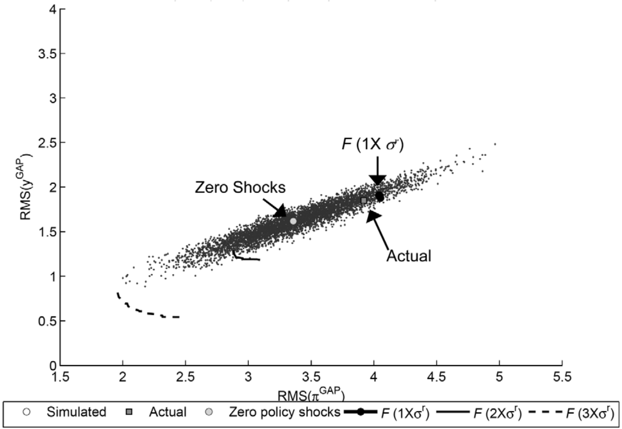
Figure 3 presents simulation results for the policy evaluation of 2003. Panel A shows 2003 to be a year in which the actual policy outcome (represented by the dark point labeled "Actual") was dominated by that of the zero-shocks counterfactual scenario (the light gray point labeled "Zero Shocks"). Thus, the shocks actually chosen by policymakers were not in the desired direction.

We also see that the actual outcome—in terms of inflation and output volatilities—is somewhat better than those consistent with the frontier that limits interest rate volatility to the sample volatility $F(1\text{X}\sigma^r)$. However, the table below panel A reveals that the within-year interest rate volatility was actually twice as much as that of the sample. When allowing for such volatility, the RMSs could be reduced, as can be learned by the location of the relevant frontier $F(2\text{X}\sigma^r)$. Allowing for even greater variability enables greater improvement in both the inflation and output gaps' RMS. However, these could only be generated by "abnormal" monetary policy shocks, as we see that the $F(3\text{X}\sigma^r)$ frontier is well outside the cluster cloud of randomly drawn monetary policy shocks.

It is also evident from the diagram that the frontiers are extremely short and that the cluster cloud is narrow and rising from left to right. This is a result of the low level of inflation throughout most of the year and the negative output gap. In such circumstances, lower interest rate paths would be more efficient, as they reduce the

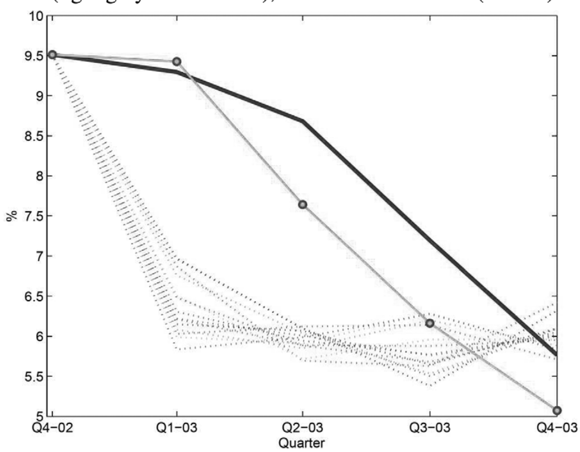
Figure 3. Ex Post Monetary Policy Evaluation for 2003

A. Actual, Counterfactual, and Frontier (F)* Root Mean Square (RMS) of Output and Inflation Gaps



	Sample Mean	Evaluated Year Actual
Within-Year S.D. of Interest Rate (σ^r)	0.71%	1.40%

B, Interst Rate Path: Actual (solid black), Zero Shocks (light gray with marker), and Frontier Paths** (dashed)



* The figure includes three frontiers where the within-year interest rate standard deviations are limited by once, twice, and three times the sample mean.

** Frontier paths are limited by twice the sample mean of the within-year interest rate standard deviation (2xσ^r).

RMSs of both inflation and output gaps at the same time. Thus, in this year the policy trade-off was between jointly stabilizing the inflation and output gaps and keeping interest rate variability low. That is, the policy dilemma was how much interest rate volatility should be allowed in the process of reducing the interest rate for the benefit of higher inflation and activity.

Panel B of figure 3 indicates that the frontier-consistent interest rate paths are such that the interest rate is sharply reduced to approximately 6 percent in the first quarter of the year and remains at this level thereafter, throughout the year. This shape of paths reaches the lowest possible (average) level of interest rate given the imposed limitation on interest rate volatility (twice the sample volatility). Note that the actual interest rate was also reduced to 6 percent, but only later, by the end of the year.

The ex post conclusion is therefore that the interest rate path in 2003 was sub-optimal. It can only be seen as inefficient ex ante as well, if the low inflation from the second quarter onwards could have been foreseen. Indeed, although inflation was brought down by large nominal exchange rate appreciations, which are in general hard to predict, inflation expectations (derived from financial market data) were already below target in the second quarter of the year, suggesting that the central bank might have started to react somewhat too late, even from an ex ante perspective. As for the degree of smoothing, the somewhat traumatic experience of the sharp interest rate cut in 2002 was certainly on the minds of policymakers. Notwithstanding, we have seen that the evaluation framework only suggests that the cuts should have been made earlier, not larger. Note that the above analysis is not intended to criticize policy decisions in 2003. We are merely demonstrating that the framework generates an ex post picture of a seemingly sub-optimal policy, which can trigger important discussions on key policymaking issues, such as the implications of interest rate smoothing and the quality of inflation forecasting.

3.3 Evaluation of 2004: Efficient Policy Shocks Generated Absolute Improvement

After the relatively dramatic macroeconomic events in 2002–3, the year 2004 brought some stability. Thus, inflation was positive, for the most part, and the negative output gap stopped widening in the second quarter. By the second quarter of the year, the interest rate had reached 4 percent, which is below the model's long-run rate. This was partly the result of negative policy shocks (see figure 2).

Figure 4 presents the policy evaluation for 2004. Panel B indicates that, similar to the case of 2003, the frontier-consistent policy paths (limited by twice the sample interest rate volatility) were generally lower than the actual one, suggesting that the reduction of the interest rate should have continued throughout the year to approximately 3 percent by the last quarter. In contrast to 2003, however, panel A shows that policy shocks in 2004 brought about improvement in the RMS of both the output and inflation gaps relative to the zero-shock alternative. Hence, this is an example of a year in which active policy decisions were in the desired direction, in terms of output and inflation stability. Yet, the frontiers demonstrate that, from an ex post point of view, greater RMS improvement could have been achieved at the cost of higher interest rate volatility.

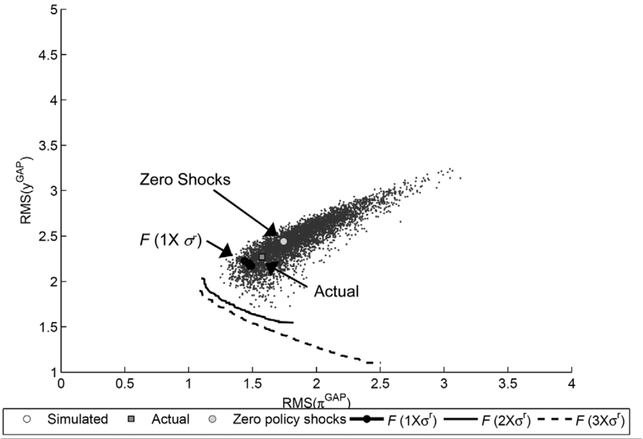
3.4 Evaluation of 2011: Trade-Off between Inflation and Output Stabilization

The year 2011 was characterized by a relatively high rate of inflation in the beginning of the year, mainly due to a large increase in housing prices, and a negative output gap in the aftermath of the Global Financial Crisis (though narrower than the previous year). Monetary policy was characterized by interest rate hikes, which seem consistent with the interest rate rule, along with some relatively small negative monetary policy shocks (see figure 2).

Figure 5 presents the results for policy evaluation of 2011. It is evident from panel A that the actual RMSs are almost identical to the zero-shocks benchmark (as expected, considering the small monetary policy shocks). It is interesting to see that all three frontiers are virtually overlapping and lie on the border of the cluster cloud of random draws. This indicates that neither "abnormal" shocks nor increased interest rate volatility were necessary in order to reduce the RMSs of inflation and output stabilization. Moreover, we can infer that the actual outcome (i.e., inflation and output RMSs) could have been improved without significantly increasing interest rate volatility (since the actual interest rate standard deviation in that year,

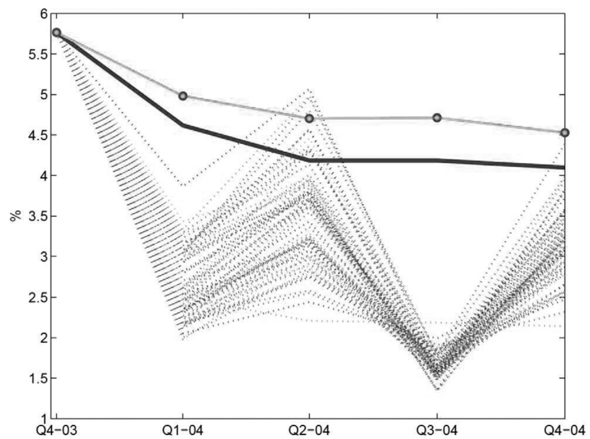
Figure 4. Ex Post Monetary Policy Evaluation for 2004

A. Actual, Counterfactual, and Frontier (F)* Root Mean Square (RMS) of Output and Inflation Gaps



	Sample Mean	Evaluated Year Actual
Within-Year S.D. of Interest Rate (σ^r)	0.71%	0.62%

B. Interest Rate Path: Actual (solid black), Zero Shocks (gray with marker), and Frontier Paths** (dashed)

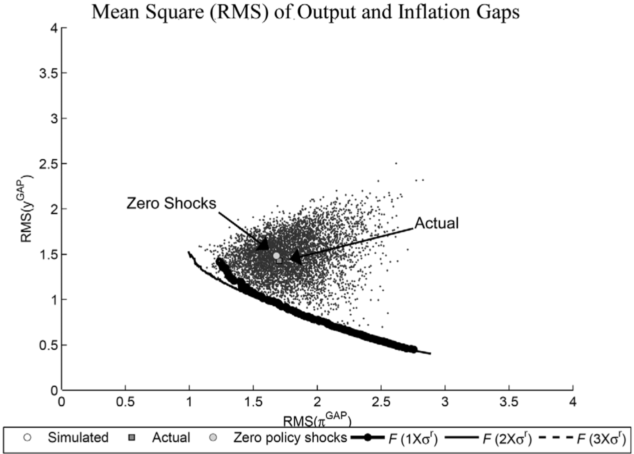


* The figure includes three frontiers where the within-year interest rate standard deviations are limited by once, twice, and three times the sample mean.

** Frontier paths are limited by twice the sample mean of the within-year interest rate standard deviation (2xσ^r).

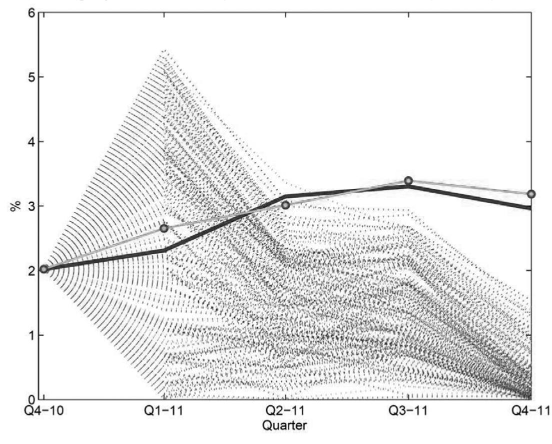
Figure 5. Ex Post Monetary Policy Evaluation for 2011

A. Actual, Counterfactual, and Frontier (*F*)* Root Mean Square (RMS) of Output and Inflation Gaps



	Sample Mean	Evaluated Year Actual
Within-Year S.D. of Interest Rate (σ^r)	0.71%	0.50%

B. Interest Rate Path: Actual (solid black), Zero Shocks (gray with marker), and Frontier Paths** (dashed)



* The figure includes three frontiers where the within-year interest rate standard deviations are limited by once, twice, and three times the sample mean.

** Frontier paths are limited by twice the sample mean of the within-year interest rate standard deviation (2xσ^r).

at 0.5 percent, was slightly lower than the sample average, 0.7 percent, for which the frontier is drawn). While there was no trade-off between the desire for a low interest rate volatility and the other policy objectives, we can see that there was a clear trade-off between inflation and output-gap stabilization (as reflected by the slope of the frontiers). From panel B we can see that, according to one tail of the distribution of efficient interest rate paths, the interest rate should have been reduced in the beginning of the year and remained low (supporting the stabilization of the output gap) and, according to the other tail, it should have been raised in the first quarter and reduced thereafter (serving the stabilization of inflation). In practice, and in accordance with the interest rate rule, the interest rate was gradually raised throughout the year.

4. Robustness Analysis

As in any econometric evaluation, there is some uncertainty concerning the results and the conclusions. In the context of the evaluation presented in the previous section, important sources of uncertainty are the actual historical shocks and the output-gap objective of the policymaker—both unobserved to the researcher. This section examines the sensitivity of the conclusions in our implementation with respect to these two sources of uncertainty.[10] In general, the results from the implementation presented above seem to be fairly robust to both sources of uncertainty.

4.1 Sensitivity to the Historical Shocks Extracted

When implementing the suggested evaluation framework, we applied the Kalman filter in order to smooth out the unobservable historical shocks. However, since the model features more shocks than observed variables, the shocks are not identified. That is, there are infinitely many combinations of shocks that are consistent with the observed variables. The Kalman filter smooths out the mean of the joint distribution of the shocks (conditional on the model, its parameters, and the full sample of the observed variables). In this section we

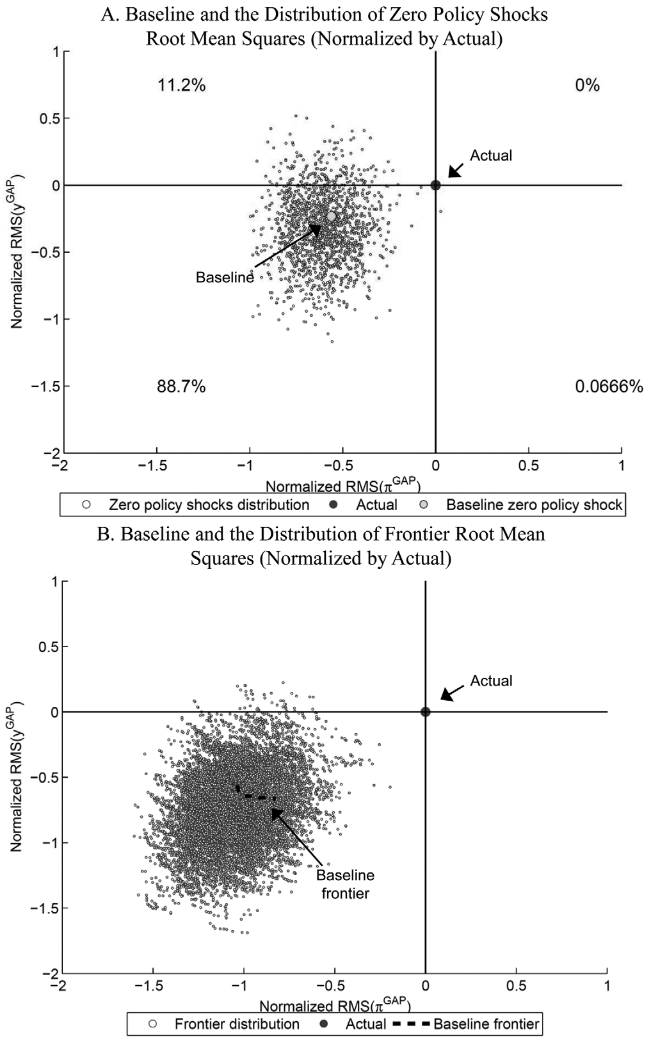
[10] The structure of the economy—namely, the model and its parameter values—is an additional source of uncertainty.

examine the sensitivity of the results of our implementation to the uncertainty presented by this distribution. To perform the test for a specific evaluated year, we first drew 1,500 sets of shocks from the conditional distribution.[11] For each set of shocks, we computed the associated RMSs of inflation and the output gap (which we refer to below as "actual"), the zero-policy-shock RMSs, and the frontier (for which we used 3,000 additional simulations). Note that although the observed variables are identical across all 1,500 iterations, the associated RMSs may nevertheless vary, as the output gap is unobservable and due to the inclusion of a forecast period in the RMS calculation (see equations (4) and (5)). Note that while the absolute location of each "actual" RMS point is uncertain, there may be less uncertainty regarding its relative location—with respect to the RMS point associated with zero policy shocks or the frontier of RMS points—which is what is relevant for our conclusions.

Figures 6–8 present results from the sensitivity analysis for the evaluated years 2003, 2004, and 2011. Each figure presents the distribution of zero-policy-shock RMSs and frontier RMSs points, all relative to the "actual" RMSs of each iteration. Thus, we "normalize" each counterfactual RMS point by subtracting from it the corresponding "actual" RMS point, so that the origin of the figure represents all the "actual" RMS points. By so shifting the location of the RMS points, we are better able to visually present our robustness analysis results, while focusing attention on the RMS points' relative location, as discussed above. To illustrate the analysis, we will discuss the robustness test for the 2004 policy evaluation depicted in figure 7. Panel A allows examining the significance of the difference between the RMSs associated with zero policy shocks and those associated with the "actual" policy shocks. For the 2004 evaluation, we find that 74.4 percent of the normalized zero-policy-shock points are in the upper-right quadrant, implying that the absolute improvement brought about by active monetary policy is fairly robust to the uncertainty associated with historical shocks. An even more robust result is that the policy shocks did not cause an absolute worsening, which is evident in the result that only 0.27 percent of the normalized

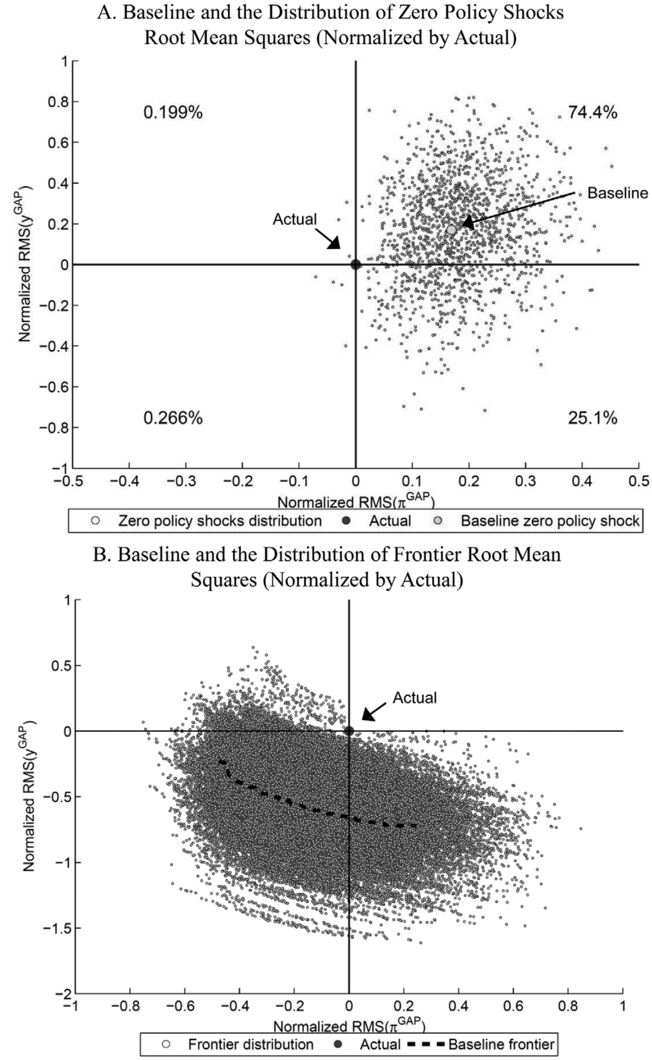
[11] The shocks are drawn using the algorithm of Waggoner and Zha (1999), where we take historical values of observed variables to be (what they refer to as) hard conditioning restrictions. Alternatively, one could use the equivalent simulation smoother of Durbin and Koopman (2002).

Figure 6. Sensitivity Analysis for 2003 Ex Post Evaluation



RMS points are located in the bottom-left quadrant. Note that we are much more confident regarding the improvement in the inflation RMS (as 99.5 percent of the points are in the right-hand half) than for the output gap (as 25.4 percent are in the lower half). We further discuss this point below. In panel B we can see the normalized

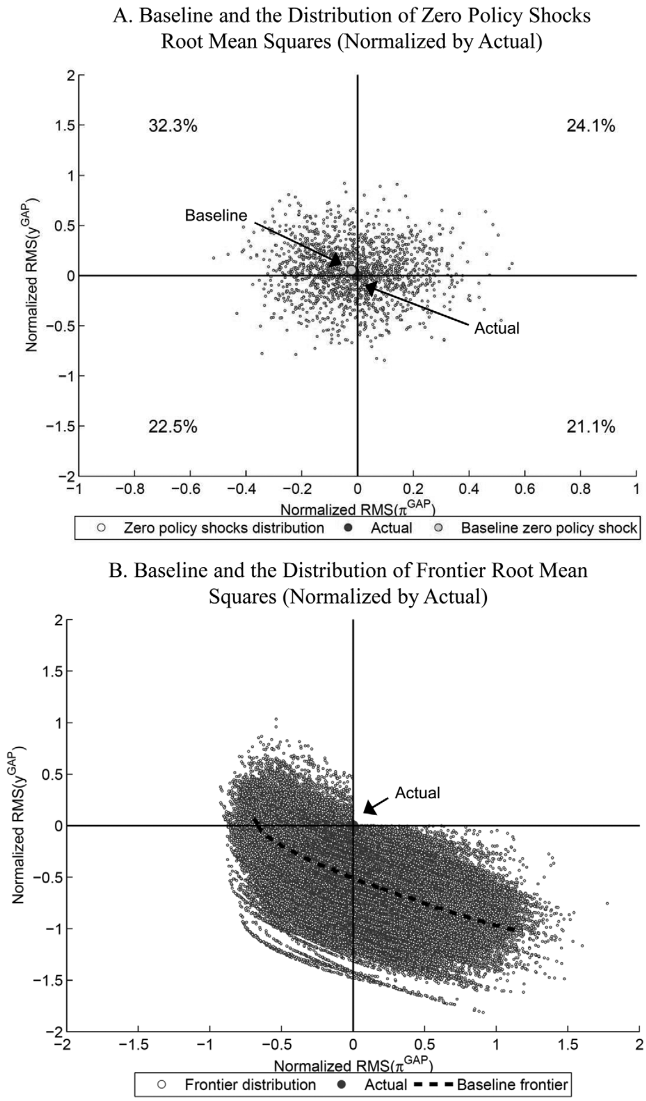
Figure 7. Sensitivity Analysis for 2004 Ex Post Evaluation



location of all $F(2\mathrm{X}\sigma^r)$ frontier points.[12] The figure helps us to assess the uncertainty regarding the distance of the "actual" RMS from the

[12]For the sake of brevity, we only report the analysis with the frontiers that limit the within-year interest rate volatility to twice the sample average.

Figure 8. Sensitivity Analysis for 2011 Ex Post Evaluation



frontier. However, unlike the case of zero policy shocks discussed above, we cannot quantify the statistical significance of the results, since any quantitative definition for "being close to the frontier" would be arbitrary. Still, the graph shows that the mass of frontier points is located some distance from the origin (that is, from the "actual" RMS). This result supports the conclusion that the chosen policy shocks in 2004 did not push the RMSs all the way down to the frontier. This conclusion is even more strongly evident in panel B of figure 6 for the year 2003. In contrast, the results for 2011 (figure 8), seem to be consistent with the hypothesis that actual RMSs were in the environment of the frontier.

After demonstrating the sensitivity analysis concerning the shocks smoothing, let us briefly review the factors that affect the magnitude of this uncertainty. The first type of factor pertains to the properties of the estimated model: under-identification of the shocks, as well as parameter uncertainty, induces an uncertainty with respect to the historical realizations of the shocks, which affect the robustness of the evaluation results.[13] The second type of factor concerns the policy objectives in the evaluation framework. The "actual" state of unobserved variables (such as the output gap) depends on the smoothed historical shocks. Therefore, the inclusion of an unobservable variable in the policy criterion increases the uncertainty surrounding the evaluation result. This is evident in panel A of figures 6 and 7: the uncertainty regarding the (qualitative) position of the zero-policy-shock point relative to the actual RMS point is concentrated on the relative location of the output-gap RMS, whereas there is almost no uncertainty with respect to the direction of change of the inflation-gap RMS. Finally, an additional factor is the inclusion of the forecast period in the RMS calculation. This introduces uncertainty concerning the location of the "actual" RMS, even if both objective variables are observable. This may increase the uncertainty regarding the relative location of RMS points. To conclude, when applying the evaluation framework—to a specific country, model, or period—it is important to keep in mind these sources of uncertainty and try to address them by the relevant sensitivity tests.

[13]Our analysis is based on deviations from an estimated policy rule. In order to address the concern that the policy rule might have changed during the sample period, we reestimated the policy rule over different subsample periods. We haven't found any evidence for substantial changes in the policy rule's parameters over time.

4.2 Sensitivity to the Definition of the Output Gap

There are various alternative definitions or measures for the unobservable "output gap." In the presented implementation, we defined the output gap as the deviation of the actual output from an unobserved, technology-driven trend (which consists of both permanent and transitory technology shocks). In this subsection we examine the robustness of our results with respect to the following alternative definitions of the output gap:

- *Baseline:* Output's deviation from a technology-driven trend, including both permanent and transitory technology shocks ($\log \frac{Y_t}{Z_t \varepsilon_t} - \log y$). This is the measure we employed for the analysis presented above.
- *Alternative Output Gap:* Output's deviation from a smoothed technology-driven trend, including only the permanent technology shocks ($\log \frac{Y_t}{Z_t} - \log y$).
- *Unemployment Gap:* This measure is derived from the way Gali, Smets, and Wouters (2012) introduce unemployment into a New Keynesian model. The unemployment gap is defined as the gap between the amount of labor input that households would like to supply under the prevailing wage and the actual amount of labor input. Since unemployment is countercyclical, we take the unemployment gap with a negative sign.

Figure 9 presents the estimated historical paths of the three alternative measures. The figure reveals that the three estimates are generally similar, with the exception of only a few periods in which the signs of the measures differ.[14] Focusing on the relevant time period of our ex post analysis (2001–11), we can see that, according to all three alternative measures, the output gap was negative most of the time, in line with the widespread view of the business cycles in Israel, as was noted in subsection 3.1.

Table 1 summarizes the results of some sensitivity tests conducted in order to assess the robustness of the main results in the

[14]Note that the sign of the output-gap measure determines whether the interest rate set by monetary policy should be higher or lower in order to improve the output gap.

Figure 9. Alternative Output Measures

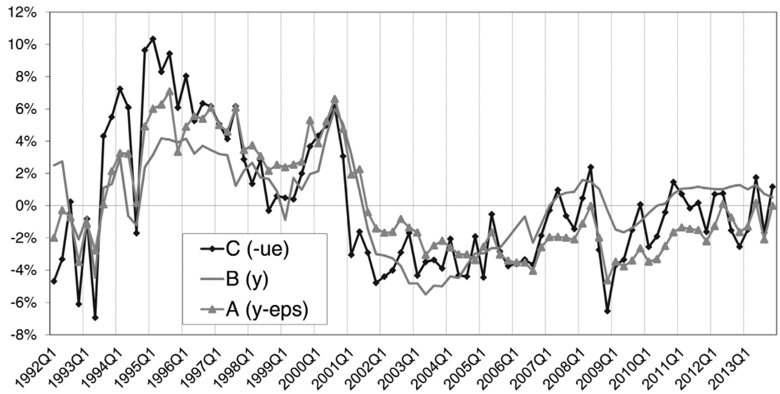


Table 1. The Main Results under Different Output Measures

	Actual-ZS			Q above Frontier			Q below Frontier		
Year	**A**	**B**	**C**	**A**	**B**	**C**	**A**	**B**	**C**
2001	?	+	+	1	3	3	0	0	1
2002	?	+	?	1	1	1	0	0	1
2003	+	+	+	3	3	3	0	0	0
2004	−	−	−	2	2	2	0	0	0
2005	?	?	?	2	1	1	0	0	0
2006	+	+	+	4	1	4	0	0	0
2007	?	?	?	1	1	1	0	0	0
2008	−	−	−	1	1	1	0	2	0
2009	?	?	?	1	0	1	1	1	1
2010	?	−	?	1	1	1	0	3	2
2011	?	?	?	2	1	1	0	1	1

implementation with respect to the different measures of the output gap. The block titled “Actual-ZS” describes the position of the actual outcome relative to the zero shocks (counterfactual) outcome, in the following way:

- The notation "–" denotes that the RMS of both the inflation and the output gap is smaller for the actual RMS outcome (i.e., the policy yielded an absolute improvement).
- The notation "+" denotes that the RMS of both the inflation and the output gap is larger for the actual RMS outcome (i.e., the policy yielded an absolute deterioration).
- The notation "?" denotes that one of the RMS indices has improved, while the other has deteriorated, so that none of the points dominates the other.

The table shows that except for the years 2001, 2002, and 2010, the position of the actual outcome relative to the zero-shocks outcome does not depend on the specific measure of the output gap.

The two other blocks in the table examine the robustness of the conclusions derived from the figures that compare the actual interest rate path with the interest rate paths on the efficient frontier.[15] The block titled "Q above Frontier" counts, for each output measure, the number of quarters during the year in which the interest rate was higher than all of the efficient interest rate paths. Similarly, the block "Q below Frontier" counts the quarters where the interest rate was lower than all of the efficient interest rate paths. In this aspect too, the table shows that despite some differences in a few years, most of the conclusions are robust to the alternative output measures.

5. Concluding Remarks

We have presented an analytical framework for conducting an ex post evaluation of monetary policy decisions. The evaluation is based on a comparison between actual and counterfactual root mean squares (RMSs) of two objective variables (typically the inflation and output gaps) during a given period. The counterfactual RMSs are generated using counterfactual interest rate paths, calculated by drawing alternative monetary policy shocks. In addition, we search for monetary policy shocks that would have generated RMSs located on the policy frontiers, that is, where it is impossible to further improve both RMSs simultaneously. In line with the view that

[15] Here we refer to the frontier that is limited to twice the sample average of within-year interest rate standard deviation $F(2\mathrm{X}\sigma^r)$.

central banks also care about interest rate smoothness, we examine different frontiers, each consistent with a different constraint on interest rate volatility.

A natural benchmark is the counterfactual case of zero policy shocks, i.e., a hypothetical case in which the central bank strictly follows the interest rate rule. Policy decisions are considered as yielding an absolute improvement (deterioration) when actual RMSs of both objective variables are lower (greater), compared with those resulting from this counterfactual benchmark scenario of zero policy shocks. We regard policy decisions as ex post efficient when actual RMSs are close to the frontier.

As an implementation, the proposed evaluation framework was applied to the Israeli economy for selected years. The results allow the categorizing of the evaluated years according to the following criteria: years characterized by RMSs which were close to the efficient frontier (e.g., 2011) or far away from it (e.g., 2003 and 2004); years where monetary policy shocks led to an absolute improvement—that is, to reduction in the RMS of both the inflation and output gaps—(e.g., 2004) or an absolute worsening (e.g., 2003); and years during which there was a policy trade-off between all three objective variables, including interest rate volatility (e.g., 2004), or within a subset of the objectives (e.g., 2003 and 2011).

The suggested framework has some clear strengths: it can be applied to many models, it provides flexibility in choosing the main policy objective variables and the period of evaluation, and it does not require the assessor to take a position with regard to the weights in the central bank's loss function. However, there are some weaknesses worth noting as well, which are important to keep in mind.

First, the results depend on the model being employed. To some extent, the issue of possible misspecification is common to almost any economic model. It is therefore preferable to apply the evaluation framework using more than one model (if possible). As noted in the introduction, to qualify for this purpose, models should include the objective variables of interest, as well as allow for the identification of shocks hitting the economy—in particular, shocks to the interest rate rule. To demonstrate an implementation of the evaluation framework, we applied it using the DSGE model employed by Bank of Israel's staff, MOISE. The paper describing this model, Argov et al. (2012), presents the model's properties and the way it interprets the data. This includes the model's impulse response functions, forecast-error-variance decomposition, autocorrelations and cross-correlations (second moments) generated by the model (compared with those observed in the data), and an evaluation of in-sample forecasting quality. It concludes that the model does a fairly good job of accounting for the economic developments in the Israeli economy during the sample period. Hence, the model has been employed at the Bank of Israel as a tool in the process of forecasting, as well as for policy evaluation using the evaluation framework suggested by the present paper.

A second issue to keep in mind is that since shocks are unpredictable, the ex post approach cannot address the question of whether policy could have been more efficient in an ex ante sense. Thus, policymakers are judged (partly) based on how they reacted (in advance) to possibly unpredictable shocks. Hence, results should be interpreted accordingly. It should also be noted that the results may be sensitive to the choice of objective variables examined.

Finally, the evaluation framework views the error term of the interest rate equation as an active monetary policy shock, a view which is open to criticism. Admittedly, the error term may also represent real-time measurement errors in the variables that drive the policy rule, or errors resulting from a misspecified rule.

All these issues should be kept in mind when employing the proposed evaluation framework, and they call for the development of additional frameworks and methodologies for policy evaluation.

We believe that an evaluation on a regular basis, say once a year, may help central banks improve their understanding of past policy decisions. It may also assist central banks in communicating their decisions, and their consequences, and identifying weaknesses in the analyses and discussions preceding policy decisions.

Appendix

Search Algorithm for Frontier Investigation

In the counterfactual simulations for an evaluated period, we repeatedly drew the four-period vector of monetary policy shocks (η^R) from a zero-mean distribution $N(0, \sigma^2_{\eta^R})$, as described in subsection 2.3 (step 4). But while searching for the feasible policy frontiers, we use the following algorithm, intended to push draws toward the frontier, thus economizing computer time:

1. Set initial benchmark values—for the mean of the policy-shock distribution to draw from, $\eta^R_{benchmark} = 0$, and for the RMSs, $RMS\left(\pi^{GAP}_{benchmark}\right) = RMS\left(\hat{y}^{GAP}_{benchmark}\right) = \infty$. Also, select the total number of counterfactual simulations, F, and initialize the simulation number, $f = 1$.

2. Draw shock values for the policy-shock vector, η^R_f, using the distribution $N(\eta^R_{benchmark}, \sigma^2_{\eta^R})$ for each element in the vector, and run a counterfactual simulation with this vector of policy shocks as described in subsection 2.3. Store the RMSs associated with this simulation as $RMS\left(\pi^{GAP}_f\right)$ and $RMS\left(\hat{y}^{GAP}_f\right)$.

3. If draw f satisfies the limit for the within-year interest rate variability, and if it falls within the desired area (defined below), update the benchmark values: $\eta^R_{benchmark} = \eta^R_f$, $RMS\left(\pi^{GAP}_{benchmark}\right) = RMS\left(\pi^{GAP}_f\right)$, and $RMS\left(\hat{y}^{GAP}_{benchmark}\right) = RMS\left(\hat{y}^{GAP}_f\right)$.

4. If $f < F$, then set $f \rightsquigarrow f+1$ and go back to step 2. Otherwise, terminate.

Definition 1. *Draw f is considered to fall within the "desired area" if at least one of the following conditions holds:*

- $RMS\left(\pi^{GAP}_f\right) < RMS\left(\pi^{GAP}_{benchmark}\right)$ *and* $RMS\left(\hat{y}^{GAP}_f\right) < RMS\left(\hat{y}^{GAP}_{benchmark}\right)$.
- $RMS\left(\pi^{GAP}_f\right) < RMS\left(\pi^{GAP}_{benchmark}\right)$ *and* $\left(RMS\left(\pi^{GAP}_{benchmark}\right) - RMS\left(\pi^{GAP}_f\right)\right) > \alpha * \left(RMS\left(\hat{y}^{GAP}_f\right) - RMS\left(\hat{y}^{GAP}_{benchmark}\right)\right)$.
- $RMS\left(\hat{y}^{GAP}_f\right) < RMS\left(\hat{y}^{GAP}_{benchmark}\right)$ *and* $\left(RMS\left(\hat{y}^{GAP}_{benchmark}\right) - RMS\left(\hat{y}^{GAP}_f\right)\right) > \left(\frac{1}{\alpha}\right) * \left(RMS\left(\pi^{GAP}_f\right) - RMS\left(\pi^{GAP}_{benchmark}\right)\right)$.

Using this algorithm, we essentially center the shocks distribution around draws that, compared with the previous draw, generate an improvement in both RMSs, or at least "a big enough" improvement in one of them. Whether an improvement is "big enough" or not depends on the parameter $\alpha > 0$, which is calibrated as the frontier slope, based on preliminary simulations.

References

Adolfson, M., S. Laséen, J. Lindé, and M. Villani. 2007. "Bayesian Estimation of an Open Economy DSGE Model with Incomplete Pass-Through." *Journal of International Economics* 72 (2): 481–511.

———. 2008. "Evaluating an Estimated New Keynesian Small Open Economy Model." *Journal of Economic Dynamics and Control* 32 (8): 2690–2721.

Argov, E., A. Barnea, A. Binyamini, E. Borenstein, D. Elkayam, and I. Rozenshtrom. 2012. "MOISE: A DSGE Model for the Israeli Economy." Discussion Paper No. 2012.06, Bank of Israel.

Beneš, J., A. Binning, M. Fukac, K. Lees, and T. Matheson. 2009. *K.I.T.T.: Kiwi Inflation Targeting Technology.* Reserve Bank of New Zealand.

Bernanke, B., and F. Mishkin. 1992. "Central Bank Behavior and the Strategy of Monetary Policy: Observations from Six Industrialized Countries." In *NBER Macroeconomics Annual 1992*, Vol. 7, ed. O. J. Blanchard and S. Fischer, 183–238. National Bureau of Economic Research, Inc.

Brubakk, L., T. Husebo, J. Maih, K. Olsen, and M. Ostnor. 2006. "Finding NEMO: Documentation of the Norwegian Economy Model." Staff Memo No. 2006/6, Norges Bank.

Christoffel, K., G. Coenen, and A. Warne. 2008. "The New Area-Wide Model of the Euro Area: A Micro-Founded Open-Economy Model for Forecasting and Policy Analysis." ECB Working Paper No. 944.

Durbin, J., and S. J. Koopman. 2002. "A Simple and Efficient Simulation Smoother for State Space Time Series Analysis." *Biometrika* 89 (3): 603–15.

Gali, J., F. Smets, and R. Wouters. 2012. "Unemployment in an Estimated New Keynesian Model." *NBER Macroeconomics*

Annual 2011, Vol. 26, ed. D. Acemoglu and M. Woodford, 329–60.

Murchison, S., and A. Rennison. 2006. "ToTEM: The Bank of Canada's New Quarterly Projection Model." Technical Report No. 97, Bank of Canada.

Seneca, M. 2010. "A DSGE Model for Iceland." Working Paper No. 50, Central Bank of Iceland.

Svensson, L. 2005. "Monetary Policy with Judgment: Forecast Targeting." Technical Report No. 819, University Library of Munich, Germany.

———. 2012. "Evaluating Monetary Policy." In *The Taylor Rule and the Transformation of Monetary Policy,* ed. E. F. Koening, R. Leeson, and G. A. Kahn, 245–74. Hoover Institution Press.

Taylor, J. B. 2011a. "Macroeconomic Lessons from the Great Deviation." In *NBER Macroeconomics Annual 2010*, Vol. 25, ed. D. Acemoglu and M. Woodford, 387–95. University of Chicago Press.

———. 2011b. "The Rules-Discretion Cycle in Monetary and Fiscal Policy." *Finnish Economic Papers* 24 (2): 78–86.

Waggoner, D. F., and T. Zha. 1999. "Conditional Forecasts in Dynamic Multivariate Models." *Review of Economics and Statistics* 81 (4): 639–51.