Assessing the Accuracy of the Aggregate Law of Motion in Models with Heterogeneous Agents

Wouter J. DEN HAAN* †

July 13, 2009

Abstract

This paper shows that the R^2 and the standard error have fatal flaws and are inadequate accuracy tests. Using data from a Krusell-Smith economy, I show that approximations for the law of motion of aggregate capital, for which the true standard deviation of aggregate capital is up to 14% (119%) higher than the implied value and which are thus clearly inaccurate, can have an R^2 as high as 0.9999 (0.99). Key in generating a more powerful test is that predictions of the aggregate law of motion are *not* updated with the aggregated simulated individual data.

Key Words: Numerical solutions, simulations, approximations

JEL Classification: C63

*Department of Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, The Netherlands and CEPR. E-mail: wdenhaan@uva.nl. Phone: +31 20 5255237. Fax: +31 20 5254254.

†I would like to thank Olivier Allais, Philip Jung, Per Krusell, Pontus Rendahl, and Tony Smith for useful comments.

1 Introduction

Models with heterogeneous agents and aggregate risk play an important role in modern macroeconomics. Most algorithms follow den Haan (1996), Krusell and Smith (1997, 1998), and Ríos-Rull (1997) and summarize the cross-sectional distribution with a finite set of moments. Krusell and Smith (1998) use the following iterative scheme. An aggregate law of motion specifies how next period's cross-sectional moments depend on current moments and aggregate productivity shocks. Taking this aggregate law as given, the algorithm solves for the individual policy rules. Using the individual policy, a panel data set is simulated for a large, but finite, number of agents. The simulated data are used to construct a time series for the cross-sectional moments. These series are then used to update the coefficients of the approximating law of motion of the aggregate variables.[1] The two accuracy tests considered in the literature to check the fit are the R-square, R^2, and the standard error of the regression equation, $\widehat{\sigma}_u$. According to the model, the fit should be perfect, that is, the R^2 should be equal to 1. The R^2 and $\widehat{\sigma}_u$ were originally proposed in Krusell and Smith (1998). Although they give most emphasis to these two tests, they considered in addition a battery of alternative tests. Papers in the literature, however, typically only consider the R^2 and $\widehat{\sigma}_u$. This is unfortunate because—as will be shown in this paper—the R^2 and $\widehat{\sigma}_u$ are very weak accuracy tests.

Simple example to document the problems. In Table 1, I report the R^2 and some properties of different aggregate laws of motion using a time series of $10,000$ observations. The aggregate capital stock, K_t, is constructed from a panel of individual data generated by the individual policy rules of Young (2009) for the model of Krusell and Smith (1998). The first row corresponds to the fitted law of motion of the regression equation:

$$\ln(K_t) = \alpha_1 + \alpha_2 a_t + \alpha_3 \ln(K_{t-1}) + u_t. \tag{1}$$

This equation has an R^2 equal to 0.99999729 and the estimated value for α_3 is equal to 0.96404. In the subsequent specifications, I change α_3 and simultaneously adjust α_1 to

[1] den Haan (1996) solves for the individual policy functions using the simulated panel, which avoids having to specify an explicit law of motion for the aggregate variables.

ensure that the mean error term of the regression equation remains equal to zero. The adjustment of α_1 also ensures that the implied mean for $\ln(K_t)$ remains the same.

As I lower the value of α_3, the value of the R^2 obviously goes down. But the changes in α_3 are such that the R^2 remains quite high. In particular, I lower the value of α_3 until the R^2 is equal to either 0.9999, 0.999, or 0.99. Despite the high R^2 values, the alternative aggregate laws of motion are very different laws of motion. This is made clear by the standard deviation of $\ln(K_t)$ that is implied by the three alternative aggregate laws of motion. The standard deviation implied by the original regression equation is equal to 0.0248, which corresponds very closely to the standard deviation of the underlying series. But as the value of α_3 is changed, the implied standard deviation plummets. For example, when α_3 is equal to 0.9324788 (0.8640985), then the true value of the standard deviation of the aggregate capital stock (the one implied by the individual policy rules) is 43% (119%) above the value implied by the approximating aggregate law of motion, even though the R^2 of the approximating laws of motion is equal to 0.999 (0.99). And when α_3 is adjusted so that the R^2 is equal to 0.9999, then there is still a 14% error for the standard deviation of aggregate capital.

Main problems of existing tests.

- An accuracy test should test whether a proposed aggregate law of motion corresponds to the aggregate law of motion that is implied by the individual policy rules. The R^2 and $\widehat{\sigma}_u$, however, only check the one-period ahead forecast error. This is an extremely limited way to check whether the approximating and the true law of motion are similar. In fact, the two laws of motion are not really compared with each other, because each period one uses as the explanatory variables the actual aggregated values of the individual choices, *not* the values generated by the approximating law of motion itself. In other words, each period the true aggregates are used to "update" or "correct" the approximating law of motion. Krusell and Smith (1996, 1998) also consider 25-year forecasts instead of 1-quarter forecasts.[2] I will show that

[2] Although this test is mentioned in the published version, actual outcomes are only reported in the

this test overcomes the updating problem of the standard test and a key element of the accuracy procedure proposed in this paper is an extension of the idea to use multi-step forecasts.

- Another problematic feature of both the R^2 and $\widehat{\sigma}_u$ is that they are averages and, thus, may hide large errors. Large errors may be problematic, even if they occur infrequently and do not affect the average. Suppose that an approximating law of motion does not predict well when the economy is in a severe recession. Even if this event is rare, it may still be important for the individual agent's policy rule, for example, for determining the amount of buffer-stock savings. Krusell and Smith (1996, 1998) seem aware of this possibility and also consider the maximum error. Unfortunately, the subsequent literature typically does not.

- The R^2 scales the errors by the variance of the dependent variable. This is not necessarily problematic, because scaling may be part of a sensible accuracy test. One runs the risk, however, that scaling makes large errors look small. Even if scaling is desirable, it depends on the problem how one should scale the residuals. The R^2 scales errors with the variance of the dependent variable, but it is not clear that this is more sensible than scaling with, for example, the mean of the dependent variable. This arbitrary aspect of the R^2 can be made clear by considering the following two regression equations that are estimated using least-squares.

$$m_{t+1} = \bar{\alpha}_0 + \bar{\alpha}_1 m_{t-1} + \bar{\alpha}_2 a_t + u_t \tag{2}$$

$$m_{t+1} - m_t = \bar{\gamma}_0 + \bar{\gamma}_1 m_{t-1} + \bar{\gamma}_2 a_t + u_t \tag{3}$$

These regression equations are identical in every aspect that matters. In particular, the estimated equations have identical implications for the prediction of m_{t+1}. The R^2, however, is not the same, because the variance of m_{t+1} is not the same as the variance of $m_{t+1} - m_t$. In this paper, I will give several examples in which the R^2 of the second equation is much lower than the R^2 of the first. But why would the R^2

working paper version.

corresponding to Equation (2) be more appropriate than the R^2 of Equation (3)? Note that $\widehat{\sigma}_u$ is not affected by this problem. In fact, the value of $\widehat{\sigma}_u$ would be identical for both specifications.

The R^2 is higher if the errors are smaller *relative* to the fluctuations in the regressand. This means that it is possible that one finds a higher R^2, i.e., an improvement in the accuracy measure, if the variance of the underlying shocks increases as is the case in Krueger and Kubler (2004).

These drawbacks of the R^2 and $\widehat{\sigma}_u$ will be documented using Monte Carlo analysis. In particular, I give several examples in which approximating aggregate laws of motion that have an R^2 in excess of 0.99 (and even some that have an R^2 above 0.9999) differ in important aspects from the true law of motion.

Alternative accuracy procedure. I propose an alternative accuracy procedure that avoids the drawbacks of the R^2 and $\widehat{\sigma}_u$. The key elements of the alternative procedure are the following. First, the test compares a long simulation obtained with the aggregate law of motion *without* updating it using the actual aggregated individual choices. This independently generated time series is then compared with the time series obtained by aggregating the cross-sectional individual values. Second, the alternative procedure focuses on maximum errors. Third, the alternative procedure stresses the importance of alternative scaling choices. Besides being too weak, accuracy tests can also be too strong in the sense that solutions that are close to the true solution in most important aspects are still rejected by the accuracy test. The procedure proposed here, therefore, also consists of several exercises to investigate how serious one should take bad outcomes of the accuracy tests.

Does the R^2 lead to the wrong answer in practice? The example above showed that the R^2 can give misleading answers about the accuracy of some alternative aggregate laws of motion, but these laws of motion were not actual numerical solutions to the Krusell-Smith model. In Section 5, I will give an example of an actual numerical procedure to solve the model in Krusell and Smith (1998) that generates an extremely high R^2, but

is clearly inaccurate in at least one dimension. Thus, the R^2 can be misleading even in evaluating the accuracy of this relatively simple model. Much more complex models have now been solved and evaluated for accuracy using only the R^2 and the standard error of the regression error.[3] Whether these numerical solutions also pass more stringent accuracy tests is an important topic for future research.

2 Accuracy tests

2.1 Generating observations

Algorithms that solve models with heterogeneous agents and aggregate uncertainty typically obtain estimates of the approximating aggregate law of motion with the following two steps. First, a time series of cross-sectional moments, m_t, is simulated. Second, a regression is used to estimate the coefficients of the approximating law of motion. The simulated moments may themselves be subject to error.[4] The consequences of these errors for the accuracy tests will be discussed in Section 2.6, but for now I simply assume that there is a true law of motion that generates a sequence of observations $\{m_{t+1}\}_{t=1}^{\bar{T}+T}$ without error according to

$$m_{t+1} = \phi(m_t, a_t), \tag{4}$$

where a_t is an observed exogenous random variable with a law of motion given by

$$a_t = \rho a_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \sigma^2\right). \tag{5}$$

The first $\bar{T}$ observations are discarded to eliminate any dependence on initial conditions and the remaining observations are used to estimate the coefficients of an approximating

[3] Papers that do use more stringent tests to check the accuracy of the aggregate law of motion are, in addition to Krusell and Smith (1996, 1998), den Haan (1997), Reiter (2002), Algan, Allais, and den Haan (2008), Reiter (2009), and Silos (2007).

[4] This is especially a concern, when the simulation is based on a finite number of agents; Algan, Allais, and den Haan (2008) show that sampling error can be substantial, even when 100,000 agents are used to contruct cross-sectional moments. Algan, Allais, and den Haan (2009) discuss several simulation techniques that avoid cross-sectional sampling variation.

law of motion, $\bar{\phi}(m_t, a_t; \bar{\alpha})$, which in the experiments is assumed to be linear. Thus,

$$m_{t+1} = \bar{\phi}(m_t, a_t; \bar{\alpha}) + u_{t+1} = \bar{\alpha}_0 + \bar{\alpha}_1 m_t + \bar{\alpha}_2 a_t + u_{t+1}, \tag{6}$$

where $\bar{\alpha}$ is the vector with the coefficients of $\bar{\phi}(\cdot)$ and u_{t+1} is a numerical error term. If the approximating law of motion is identical to the true law of motion, then $u_{t+1} \equiv 0$ for every t.

2.2 Popular tests used in the literature

The R-square. The R-square, R^2, is defined as

$$R^2 = 1 - \frac{\sum_{t=\bar{T}+1}^{\bar{T}+T} \widehat{u}_{t+1}^2}{\sum_{t=\bar{T}+1}^{\bar{T}+T} (m_{t+1} - \widehat{\mu}_m)^2} \tag{7}$$

with

$$\widehat{u}_{t+1} = m_{t+1} - \widehat{m}_{t+1}, \tag{8}$$

$$\widehat{m}_{t+1} = \bar{\phi}(m_t, a_t; \widehat{\bar{\alpha}}) = \widehat{\bar{\alpha}}_0 + \widehat{\bar{\alpha}}_1 m_t + \widehat{\bar{\alpha}}_2 a_t, \tag{9}$$

and

$$\widehat{\mu}_m = \sum_{t=\bar{T}+1}^{\bar{T}+T} m_{t+1}/T. \tag{10}$$

The standard error of the regression equation. The standard error of the regression equation, $\widehat{\sigma}_u$, is given by

$$\widehat{\sigma}_u = \left(\sum_{t=\bar{T}+1}^{\bar{T}+T} \widehat{u}_{t+1}^2/T \right)^{1/2}. \tag{11}$$

This measure is sensitive to rescaling the variables or measuring them in different units. Typically, this is not a problem, since the variables are in logs and calibrated to have a plausible variance. The error term, u_t, then gives a percentage error and the value of $\widehat{\sigma}_u$ has a sensible interpretation.

2.3 Problems with existing tests

Improper correcting/updating when evaluating the aggregate law of motion. The objective is to evaluate how close the approximate law of motion $\bar{\phi}(m_t, a_t; \widehat{\bar{\alpha}})$ is to the true law of motion $\phi(m_t, a_t)$. The R^2 and $\widehat{\sigma}_u$ are weak tests for this comparison; in assessing the fit of $\bar{\phi}(m_t, a_t; \widehat{\bar{\alpha}})$ they use as the explanatory variable the values of m_t that are generated by the *true* law of motion $\phi(m_t, a_t)$. That is, observations generated with the true law of motion are used to correct and update the predictions of the approximating law of motion. If the approximating law of motion is pushing the observations in the wrong direction, then it is only allowed to do so for one period; in the next period, the true cross-sectional moments are used to put the approximating law of motion back on track. Especially if the error made by the approximating law of motion is systematic, then this could be a serious problem. This problem indicates that $\widehat{u}_{t+1}$ is not the right error term to consider.

Weak metric. The next problem of standard accuracy tests is that they do not use an appropriate metric to evaluate whether the error terms are close to zero. There are two problems. First, the R^2 and the $\sigma_{\widehat{u}}$ measure are averages and second, they can hide serious errors because of scaling.

- **Averaging.** A typical way to report errors in the numerical literature is to report *maximum* errors. Especially for this type of accuracy test it is important to use stringent standards? The reason is that the aggregate law of motion is only one component of the numerical solution of the model, and errors in the different components of the numerical solution can reinforce each other.

- **Scaling**. The R^2 scales the error term with the variance of the dependent variable. This property makes it possible to generate high values for the R^2 even though $\widehat{u}_{t+1}$ takes on problematical values. Scaling of errors is not necessarily bad, but it often is not clear what the appropriate scaling is. How arbitrary and important the choice of scaling is, can be is made clear by considering the regression equation that uses

$m_{t+1} - m_t$ instead of m_t as the dependent variable. That is,

$$m_{t+1} - m_t = \bar{\alpha}_0 + (\bar{\alpha}_1 - 1)m_t + \bar{\alpha}_2 a_t + u_{t+1}. \tag{12}$$

I will refer to this specification as the "first-difference specification" and to the specification in Equation (6) as the "level specification". The R^2 of the first-difference specification will be different from the R^2 of the level specification, even though the estimated residuals and predictions for m_{t+1} are identical. Below, I will show that the R^2 of the first-difference specification can be substantially different from the level specification; $\widehat{\sigma}_u$ does not suffer from this problem.

Overfitting. The R^2 can only increase (and $\widehat{\sigma}_u$ only decrease) if additional explanatory variables are added. But if these additional variables do not belong in the law of motion, then they only worsen the accuracy of the approximating law of motion.[5]

When is an R^2 too low? The values of the R^2 reported in Krusell and Smith (1998) are equal to or higher than 0.999985. These are indeed high values. But in Section 4.2, I will give an example in which the R^2s are above 0.9999 and the approximating law of motion is clearly missing some key features of the true law of motion. Part of the problem in interpreting the R^2 is that it takes the square. Since an accurate model has $u_{t+1} \equiv 0$ for every t, it better be the case that we are talking about small errors. By taking the square, the R^2 pushes the measure closer to 1. Suppose that the R^2 is equal to 0.99. One might think, that this indicates that the fit is quite good. But an R^2 equal to 0.99 implies that the Residual Sum of Squares is equal to 1% of the variance of m_{t+1}. That is, the standard error, $\widehat{\sigma}_u$, is equal to 10% of the standard deviation of m_{t+1}. As an empirical fit, this may be quite good, but as a numerical error this is very high.

[5] Most papers use only a few explanatory variables and a large number of observations in which case overfitting is not an issue. But this is not true for all papers, For example, Storesletten, Telmer, and Yaron (2007) use 19 moments as explanatory variables and 20,000 observations. The degrees of freedom are still high, but how many more moments could one add? The accuracy procedure described below avoids the problem of overfitting.

From the literature, it is not clear what values for the R^2 are considered to be too low. Gomes and Michaelides (2008) are happy that all their R^2s are larger than 0.99. Young (2005) argues that an R^2 equal to 0.975 is *"fairly low from a numerical standpoint"*. This makes sense, since an R^2 equal to 0.975 corresponds with a standard deviation of the regression residual that is equal to 16% of the standard deviation of the dependent variable. Nakajima (2007) finds some values for the R^2 that are (just) below 0.975, but this does not seem to be a reason for concern. Below, I show that even much higher values of the R^2 can be consistent with approximating laws of motion that differ in key aspects with the true law of motion. This difficulty in interpreting the R^2 and the problem of overfitting are also recognized as problems of the R^2 in the econometrics literature.[6]

Evaluating an R^2 is especially problematic in AR(1) type processes, which are common in macro economic models. If $\bar{\alpha}_2$ is set equal to zero in Equation (6), then the R^2 is equal to $\bar{\alpha}_1^2$ independent of the value of $\widehat{\sigma}_u$, that is, the R^2 provides no information about the goodness of fit at all. Also, under the typical case in which $0 < \widehat{\bar{\alpha}}_1 < 1$ and $\widehat{\bar{\alpha}}_2 > 0$, then the R^2 is bounded below by $\bar{\alpha}_1^2$.

The problem of figuring out what level of inaccuracy is acceptable is a problem that affects other inaccuracy tests to some extent as well.[7] The R^2 enhances the problem, because of the scaling with a volatility measure and because it is a measure of the average error.

2.4 New accuracy procedure

In this section, I propose an alternative procedure that is easy to implement, but avoids the drawbacks of the existing tests. It consists of three parts, that will be explained in

[6] Verbeek (2000) points out that (*i*) least-squares by definition leads to the highest possible R^2, but that this doesn't mean that the estimated model also has the best statistical properties and (*ii*) there is no absolute benchmark to say that an R^2 is 'high' or 'low'.

[7] This is not always the case. Santos (2000) relate the Euler equation residual to errors in the policy function. Reiter (2001) and Santos and Peralta-Alva (2005) construct a relationship between the size of the errors of the accuracy test and an *upperbound* on the error of objects economists could be interested in, such as the obtained utility level or moments. More information is given in Appendix B.

the following subsections.

I. Calculate maximum errors using observations generated without updating. As pointed out above, a problem of the standard procedure is that observations generated using the true law of motion $\phi(m_t, a_t)$ are used as the explanatory variable to predict m_{t+1}. Instead, I propose to do the following.

1. Obtain a new draw for $\{a_t\}_{t=1}^{\bar{T}+T}$. That is, the accuracy test should not be done with the sequence used to obtain the estimate $\widehat{\widehat{\alpha}}$. If T is very high and the number of regressors low, then using a new draw is unlikely to make a difference. Nevertheless, there is no excuse to use the old draw just because there are circumstances in which it doesn't matter.

2. Use the new sequence for a_t to generate a new time series for m_{t+1}. Typically, this is done by simulating a panel of individual variables and then taking the cross-sectional moments each period. In the notation of this paper, this means iterating using

$$m_{t+1} = \phi(m_t, a_t).$$

3. Let $\widehat{\widehat{m}}_{t+1}$ stand for the corresponding realization according to the approximating law of motion. It is generated as follows.

$$\widehat{\widehat{m}}_1 = m_1 \quad \text{and}$$

$$\widehat{\widehat{m}}_{t+1} = \bar{\phi}(\widehat{\widehat{m}}_t, a_t; \widehat{\widehat{\alpha}}) = \widehat{\widehat{\alpha}}_0 + \widehat{\widehat{\alpha}}_1 \widehat{\widehat{m}}_t + \widehat{\widehat{\alpha}}_2 a_t \quad \text{for } t \geq 1. \tag{13}$$

Note that $\widehat{\widehat{m}}_{t+1}$ is generated using only the approximate law of motion and *never* relies on the true law of motion ϕ or any values generated with ϕ. That is, $\widehat{\widehat{m}}_t$ is used as the explanatory variable not m_t.[8]

[8] The following notation is used throughout this paper. A bar above a symbol indicates that it is related to the approximating law of motion. A hat above a coefficient indicates the least-squares estimate and a hat above a variable indicates that it is a fitted variable using m_t as the explanatory variable, i.e., the fitted variable according to equation (9). A double hat above a variable indicates that it is a fitted variable using $\widehat{\widehat{m}}_t$ as the explanatory variable, i.e., the fitted variable according to equation (13).

The key difference between this and the standard procedure is that this procedure also checks whether errors accumulate. This is important in dynamic numerical problems. Today's decision depends on beliefs about what the world looks like tomorrow, which in turn depends on beliefs about what the world looks like the period after. Small errors in beliefs can, thus, accumulate just like small deviations between $\bar{\phi}(m_t, a_t; \widehat{\widehat{\alpha}})$ and $\phi(m_t, a_t)$ can accumulate.

I assume that m_t corresponds to the log of a variable, so that u_{t+1} is a percentage error. The simulated (percentage) error, $\widehat{\widehat{u}}_{t+1}$, is defined as[9]

$$\widehat{\widehat{u}}_{t+1} = \left| \widehat{\widehat{m}}_{t+1} - m_{t+1} \right|. \tag{14}$$

The maximum simulated percentage error is defined as

$$\widehat{\widehat{u}}^{\max} = \left\{ \widehat{\widehat{u}}_{t^*+1} : \widehat{\widehat{u}}_{t^*+1} \ \geq \widehat{\widehat{u}}_{t+1} \ \forall t \right\}. \tag{15}$$

This statistic should be the main focus of the accuracy test, although useful information may also be given by the average error:

$$\widehat{\widehat{u}}^{\text{ave}} = \frac{\sum_{\bar{T}+1}^{\bar{T}+T} \widehat{\widehat{u}}_{t+1}}{T}. \tag{16}$$

II. Investigate the "fundamental accuracy plot". Plot the time paths of m_{t+1} and $\widehat{\widehat{m}}_{t+1}$ in one graph. This graph effectively reveals important information. First, it indicates when the large errors occur. Second, it indicates whether maximum errors are large relative to average errors. Third, it indicates whether errors are persistent and/or typically have the same sign. Finally, it is often helpful in determining whether inaccuracies found matter or not. Below I will give several examples to illustrate these claims.

III. Compare properties of true and approximating law of motion. In practice, one does not know the functional form of the true law of motion, but one can generate observations for the cross-sectional moments by simulating a panel. This means one can

[9] If u_{t+1} is not expressed as a standard error, then one could use $\widehat{\widehat{u}}_{t+1} = \left| \widehat{\widehat{m}}_{t+1} - m_{t+1} \right| / m_{t+1}$, or if m_t takes on values close to zero, then one could use $\widehat{\widehat{u}}_{t+1} = \left| \widehat{\widehat{m}}_{t+1} - m_{t+1} \right| / \left(\sum_t m_{t+1}/T \right)$.

calculate, for example, impulse response functions for m_{t+1}. These can then be compared with the impulse response functions implied by the approximating law of motion $\bar{\phi}(\cdot;\widehat{\bar{\alpha}})$. This is another version of the "fundamental accuracy" plot, but now for a particular sequence of realizations of the exogenous driving process. The fundamental accuracy plot and the impulse response function check the accuracy of the approximating law of motion for typical events. One could repeat the procedure for less likely realizations of the exogenous driving process. This way one checks the accuracy of the approximating law of motion in parts of the state space that are less likely.

One can also compare the moments of m_{t+1} and $\widehat{\widehat{m}}_{t+1}$. The moments of $\widehat{\widehat{m}}_{t+1}$ can always be calculated from the simulated sequence, but if the approximating law is simple enough, then one could also calculate moments directly from the approximating law of motion itself.

Checking for accuracy should be more than simply reporting the results of some accuracy tests. It is important to "play around" with the solution obtained to make sure one understands what its properties are. Krusell and Smith (2006) propose to do this by considering the effects of changes in the initial cross-sectional distribution in a two-period model. Since the computational complexity is much smaller in this environment, it is easier to understand the properties of the model.

2.5 Discussion

In this section, I discuss the advantages of the new accuracy procedure and explain why it avoids the problems of the standard procedure. A key aspect of the new procedure is the comparison of the law of motion one wants to approximate with a time series that is generated using only the approximating law of motion. By not using observations of m_t to update the approximating time path, one obtains a much better comparison of the two laws of motion.

The metric used is the maximum error and the error is a percentage error, that is, it is scaled with the level of m_{t+1}. Note that the maximum error depends on the sample length. That is, getting a low maximum error using a long series for $\widehat{\widehat{u}}_{t+1}$ is more impressive than

getting the same low maximum error with a shorter sample. The reasons for using the maximum error are the following. First, accuracy tests should reveal the weaknesses of a numerical solution and should not hide them by averaging. Second, since inaccuracies of the different elements of the algorithm can reinforce each other, it is fundamental to impose tough standard on the different parts. The third reason is that maximum errors have been shown to be able to bound errors on implied moments, which is the kind of statistic one typically would like the approximating law of motion to predict accurately.[10]

The question arises what value for $\widehat{\widehat{u}}^{\max}$ is too high. It would be convenient if there is a magic number that specifies the lowest acceptable number for $\widehat{\widehat{u}}^{\max}$. Suppose that the maximum error in a sample of 3,000 observation is less than 0.01%. Based on my own experience, I would find it very surprising, if any property of the model is then inaccurately measured, except possibly some exotic ones. Often one does not find numbers this small for complex problems. If one does not find a small value for $\widehat{\widehat{u}}^{\max}$, then this does not mean that one should automatically discard the approximating law of motion. Many properties of the approximating law of motion may still be measured accurately. The fundamental accuracy plot provides information about what type of error occurs when. The researcher should think through, how these errors could matter. For example, if one finds that the approximating law of motion underpredicts aggregate capital during severe recessions, then one could investigate whether the corresponding errors for market prices affect individual choices.

Multi-step accuracy test. The accuracy test proposed has similarities with one of the additional tests considered in Krusell and Smith (1996, 1998), namely the one that uses 100-period ahead instead of 1-period ahead forecast errors. This error is defined as follows:

$$\begin{aligned} \widetilde{u}_{t,t+100} &= m_{t+100} - \widetilde{m}_{t,t+100}, \\ \widetilde{m}_{t,t+j+1} &= \widehat{\widehat{\alpha}}_0 + \widehat{\widehat{\alpha}}_1 \widetilde{m}_{t,t+j} + \widehat{\widehat{\alpha}}_2 a_t \quad \text{for } 0 \leq j \leq 99, \text{ and} \\ \widetilde{m}_{t,t} &= m_t \quad \text{for } t \geq 1. \end{aligned} \tag{17}$$

[10] This is shown in Santos and Peralta-Alva (2005) and discussed in appendix B.

By extending the forecast horizon from 1 to 100 periods, one does not allow the actual values of the aggregated individual observations to update the proposed aggregate law of motion during the forecasting interval. Since $\widetilde{m}_{t,t} = m_t$, however, one still uses the actual aggregated individual choices to initialize each 100-period ahead forecast. The test proposed in this paper allows for no updating at all. It uses the approximating aggregate law of motion to generate a long time series completely independently of the aggregated individual observations (the true law of motion). An important reason to consider multi-step forecast errors is that tiny errors can accumulate over time.[11] This suggests that it is more important that the forecast horizon is driven to a high number, then that the exercise is repeated many times for a fixed forecast horizon. In the examples considered in this paper, however, the 100-period ahead forecast error considered by Krusell and Smith (1996, 1998), is as powerful as the test I propose.[12]

2.6 What if m_t cannot be generated without error?

Time series for m_t are typically obtained from a simulated panel with a finite number of agents. Even with a large number of agents, the cross-sectional moments are subject to some sampling variation, which means that the law of large numbers no longer applies. That is, there is no law of motion $\phi(m_t, a_t)$ to describe the simulated m_t for which the prediction errors are exactly equal to zero. It is then possible that the generated values for $\widehat{\widehat{m}}_{t+1}$ are more accurate than the generated values m_{t+1}, because $\widehat{\widehat{m}}_{t+1}$ is not subject to cross-sectional sampling variation. That is, one could get high values for $\widehat{\widehat{u}}^{\max}$, even if the approximating law of motion is accurate. I found the following version of the "fundamental accuracy" graph helpful in determining whether inaccuracies in the generation of m_{t+1} are behind high values of $\widehat{\widehat{u}}^{\max}$. Start by choosing some deterministic time paths for the exogenous driving process. For example, if the driving process is as in Equation

[11] For example, consider $x_t = 1 + 0.99x_{t-1}$ and $x_t = 1 + 0.991x_{t-1}$, both starting at $x_1 = 1$. These processes will track each other quite closely for a long time even though their limiting values differ 11% from each other.

[12] Krusell and Smith (1996) also report the correlation between the 100-period ahead forecast and the realized outcome, but this test has very weak power.

(5) then one could consider the time path implied by setting $\varepsilon_t = 1$ for the first 25 periods, setting $\varepsilon_t = -1$ for the next 25 periods, setting $\varepsilon_t = 1$ for the next 25 periods and so on. Next, plot the corresponding time paths for m_{t+1} and $\widehat{\widehat{m}}_{t+1}$. If one finds that m_{t+1} follows the trajectory of $\widehat{\widehat{m}}_{t+1}$ closely, but that it wiggles around $\widehat{\widehat{m}}_{t+1}$, then one can be fairly certain that $\widehat{\widehat{m}}_{t+1}$ is actually more accurate then m_{t+1} and that $\widehat{\widehat{u}}^{\max}$ overestimates the inaccuracies of the approximating law of motion. The reason is that for such a deterministic time path the wiggles cannot be due to randomness of ε_t, but are most likely due to sampling error in measuring m_t.

3 Design of two experiments

3.1 Specification of a true law of motion

The reason we need accuracy tests in DSGE models is that we do not know the true law of motion. Of course, this does not make the evaluation of accuracy tests any easier. To evaluate existing and proposed tests, I simply specify a true law of motion $\phi(\cdot)$, which makes the evaluation of the accuracy of the approximating law of motion unambiguous. I chose specific specifications for $\phi(\cdot)$ to highlight the weaknesses of existing accuracy tests. Is this fair? Sure it is. All specified processes are well behaved driving processes and differ from the approximating law of motion in a sensible way. A good accuracy test should be capable of detecting the differences between the laws of motion considered and the simpler approximating law of motion. Some readers may believe that the only inaccuracies they will encounter are those that can be discovered by existing accuracy tests; readers who have so much knowledge about the true solution do not need the help of any accuracy test.

In the first experiment, the function $\phi(\cdot)$ is linear and contains both m_t and m_{t-1}, whereas the approximating law of motion only contains m_t. In the second experiment, $\phi(\cdot)$ is non-linear whereas the approximating law of motion is linear.

3.2 Steps of the Monte Carlo experiments

Phase 1: Estimate the approximating law of motion. Using the true law of motion for m_t, $\overline{T} + T$ observations for m_t are generated. The last T observations are used to estimate the coefficients of the approximating law of motion. I set $\bar{T} = 500$ and T takes on the values $3,000$ and $50,000$. There are 100 repetitions.

Phase 2: Traditional accuracy tests. In this step, I calculate the R^2 and $\widehat{\sigma}_u$ statistics.

Phase 3: New accuracy procedure. In this step, I perform the three steps of the new accuracy procedure.

- Draw a new realization for a_t and recalculate a time series for m_{t+1} using the true law of motion. Next, calculate $\widehat{\widehat{m}}_t$, without using any realization of m_{t+1}, except for m_1. Finally, calculate $\widehat{\widehat{u}}^{\max}$ and $\widehat{\widehat{u}}^{\text{ave}}$.
- Draw the fundamental accuracy plot.
- Compare properties of the true and the approximating law of motion.

Monte Carlo studies tend to make tedious reading. To reduce the burden on the reader, I report only the key results in the main text and the reader can find detailed results and additional exercises in Appendix A.

4 Results for the two experiments

4.1 Experiment 1: Missing second-order lag

4.1.1 Experiment 1: Specification

In the first Monte Carlo experiment, the true law of motion is given by

$$m_{t+1} = \alpha_0 + \alpha_1 m_t + \alpha_2 a_t + \alpha_3 m_{t-1}, \tag{18}$$

where a_t is an observed exogenous random variable with a law of motion given by Equation (5). I consider two sets of parameter values and they are reported in Table 2. Both

parameter sets imply that the impulse response function is hump-shaped. For the parameters of Experiment 1.1, the hump of the impulse response function is small, but for the parameters of Experiment 1.2, it is more substantial.

As mentioned above, it is not clear whether the errors from the accuracy tests should be scaled and—if they should be scaled—whether they should be scaled by the standard deviation of m_t, the non-stochastic mean of m_t, or something else. The answer depends on the actual context. I assume that m_t stands for the log of a variable. Unscaled errors, thus, correspond to percentage errors. Using unscaled errors means that the magnitudes of the errors increase with the variance of the driving process. Consequently, a sensible value for the variance of the driving process must be chosen. In each experiment, I choose the standard deviation of ε_t such that the standard deviation of m_t is equal to 0.025 (i.e., 2.5%) which is roughly equal to the standard deviation of log aggregate capital in the model with heterogeneous agents solved in Algan, Allais, and den Haan (2008). Occasionally, I report errors as a fraction of the standard deviation of m_t, but since this standard deviation is less than 1, this means that scaled errors are substantially larger.

4.1.2 Experiment 1: Traditional accuracy test outcomes

Table 3 summarizes the results of the traditional accuracy tests across Monte Carlo replications. The sample size used makes little difference. Thus, without loss of generality, I focus on the accuracy tests obtained with $T = 3,000$. First, consider Experiment 1.1. The R^2 is very high across Monte Carlo replications for the level regression. In particular, the minimum R^2 is equal to 0.9995. The maximum (average) value for $\widehat{\sigma}_u$ is equal to 0.049% (0.047%). A standard error of 0.049% corresponds to 1.96% of the standard deviation of m_t, which does not sound impressive at all, but is consistent with the high R^2 value.[13] For the first-difference regressions, the R^2s are a bit lower, but the minimum value across Monte Carlo replications is still above 0.99. For Experiment 1.2, the R^2 for the level equations are fairly high. In particular, the minimum (average) R^2 is equal to 0.9940 (0.9952).

[13] An R^2 equal to 0.9995 corresponds with a standard deviation of the residual equal to 2.2% of the sample standard deviation of m_t.

Interestingly, the R^2 of the first-difference equation is substantially lower. In particular, the minimum (average) value is equal to 0.8413 (0.8408). Also, the outcomes for $\widehat{\sigma}_u$ are not as low as for Experiment 1.1. The maximum (average) value of $\widehat{\sigma}_u$ across Monte Carlo replications is now equal to 0.174% (0.168%).

4.1.3 Experiment 1: New accuracy procedure

New accuracy procedure for Experiment 1.1: I - test outcomes. Table 4 reports the results for the new accuracy tests. The sample size used affects the outcomes of the test statistics somewhat, but both sample sizes give a similar picture about accuracy.[14] Therefore, I only discuss the case with $T = 3,000$ in the text. For Experiment 1.1, the average (median) across Monte Carlo replications of the maximum residual, $\widehat{\widehat{u}}^{\max}$, is equal to 0.83% (0.82%). These numbers clearly indicate that the fit of the approximating law of motion is not as good as the high R^2s suggest. These are the kind of numbers, that may lead a researcher to still accept the solution, but only after a more careful analysis. The average (median) across Monte Carlo replications of $\widehat{\widehat{u}}^{\text{ave}}$ is 0.21% (0.16%). Also not a spectacular outcome.

For Experiment 1.2, the new accuracy tests make very clear, that the approximating law of motion is not accurate. For example, the maximum error in the simulated series is on average 3.34%, which exceeds the standard deviation of m_t by far and even the minimum across Monte Carlo replications is equal to 2.48%. Such high values clearly suggests that the approximating law of motion differs in a substantial way from the true law of motion, in sharp contrast to what is suggested by the high values of the R^2 statistic. Krusell and Smith (1996) consider two statistics to evaluate 100-period ahead forecast errors. The first is the correlation between the forecast, $\widetilde{m}_{t,t+100}$, and the realization, m_{t+100}. The second

[14] In this Monte Carlo experiment, I set the length of the sample used to estimate the coefficients of the approximating law of motion equal to the length of the sample to do the accuracy test, but these could in principle be different. Given a value for $\widehat{\widehat{\alpha}}$, a longer sample in the accuracy test can only lead to an increase in $\widehat{\widehat{u}}^{\max}$. By also using a longer sample in the regression analysis, however, one may obtain a more precise aggregate law of motion and obtain a lower value for $\widehat{\widehat{u}}^{\max}$. Thus, there is not necessarily a monotone relationship and there is none found in the results.

is the maximum forecast errors, $\widetilde{u}_{100}^{\max} = \max_{\{t \leq T-100\}} \widetilde{u}_{t,t+100}$. The results are reported in Table 5. The R^2s for the comparison of $\widetilde{m}_{t,t+100}$ and m_{t+100} are low for Experiment 1.2, but the minimum R^2 across Monte Carlo replications is above 0.9929 for Experiment 1.1. The average forecast error is also not very powerful in detecting inaccuracies. The statistic $\widetilde{u}_{100}^{\max}$, however, clearly indicates that the solutions are not accurate and that the high R^2 values give a very misleading picture about the accuracy of the approximations. In fact, the behavior of $\widetilde{u}_{100}^{\max}$ is very similar to the behavior of $\widehat{\widehat{u}}^{\max}$.

New accuracy procedure for Experiment 1: II - fundamental accuracy plot. Figures 1 and 2 give a graphical presentation of the traditional and the new accuracy tests, respectively. To be precise, Figure 1 plots the outcome of the approximating law of motion calculated with

$$\widehat{m}_{t+1} = \widehat{\widehat{\alpha}}_0 + \widehat{\widehat{\alpha}}_1 m_t + \widehat{\widehat{\alpha}}_2 a_t \tag{19}$$

together with the true realizations for m_t. Figure 2 plots the outcome of the approximating law of motion calculated with

$$\widehat{\widehat{m}}_{t+1} = \widehat{\widehat{\alpha}}_0 + \widehat{\widehat{\alpha}}_1 \widehat{\widehat{m}}_t + \widehat{\widehat{\alpha}}_2 a_t, \tag{20}$$

also together with the true realizations for m_t. The initial value, m_1, is the same. Panel A of each figure reports the results for Experiment 1.1 and Panel B the results for Experiment 1.2. For both experiments, I plot the first 250 observations (after disregarding the first 500) in the first Monte-Carlo replication. In practice, one would of course want to take a look at the whole sequence, but for this experiment it does not matter very much which part of the sample one looks at and the graphs are more clear if a shorter sample is used.

For Experiment 1.1, the R^2 in the first Monte Carlo replication is equal to 0.9996 and the true and fitted values are almost always indistinguishable to the naked eye. For Experiment 1.2, the R^2 is equal to 0.9953 and although tiny errors are occasionally visible, the fit seems excellent.

Figure 2 makes clear that this is not the case when the observations of the lagged explanatory variable are generated by the approximating law of motion. Panel A documents that for Experiment 1.1, the differences are typically still small, but they occur

much more frequently than in figure 1. The maximum error in this subsample of 250 observations is 0.47%, whereas the maximum value $\widehat{\widehat{u}}^{\max}$ across Monte Carlo replications is equal to 1.18%. So by taking simply an arbitrary subsample, I do not fully reveal the existing differences between m_{t+1} and $\widehat{\widehat{m}}_{t+1}$.

Panel B of figure 2 shows the comparison of the separately generated series for Experiment 1.2. Whereas the results for Experiment 1.1 started to show some cracks, but seemed (at least in this subsample) reasonable, the results for Experiment 1.2 are clearly bad. There are long periods where the independently fitted values are quite different from the true values, indicating that the approximating law of motion is not accurate. Nevertheless, the R^2 for Experiment 1.2 in this Monte Carlo replication is equal 0.9953, not that much lower than the R^2 of Experiment 1.1.[15]

The question arises whether the new and more powerful accuracy test picks up something important. This is of course difficult to answer without knowing what the researcher is interested in. Consider Experiment 1. Figure 2.A makes clear that the approximating linear law is not as accurate as the high R^2 suggests, but the path calculated with the approximating law of motion does roughly follow the movements in the time path generated with the true law of motion. To see whether errors matter or not, one should investigate the properties of the true and approximating law more closely, which is the last step of the new accuracy procedure.

New accuracy procedure for Experiment 1: III - evaluation. How different are the true and the approximating law of motion? To answer this question, I look at a set of moments and impulse response functions. The differences in the moments and detailed summary statistics of the Monte Carlo experiment are given in Appendix A. Here I simply compare the true impulse response function with the impulse response function implied by the approximating law of motion of the first Monte Carlo replication. This summarizes in a concise manner what is going on.

[15] If one—only out of curiousity of course—would use the R^2 to evaluate the fit between m_{t+1} and $\widehat{\widehat{m}}_{t+1}$, then one would get for the data in this first Monte Carlo replication an R^2 of 0.84, substantially below 0.9953.

The two impulse response functions are reported in Figure 3. Panel A reports the results for Experiment 1.1 and Panel B for Experiment 1.2. The values of the R^2 are equal to 0.9996 and 0.9953 for Experiment 1.1 and 1.2, respectively. The graphs of the impulse response functions make clear, that the two laws are very different and that the R^2 clearly is an inadequate accuracy test.

4.2 Experiment 2: Missing non-linearity

4.2.1 Experiment 2: Specification

In the second Monte Carlo experiment, the true law of motion is given by

$$m_{t+1} = \alpha_0 + \alpha_{1,t} m_t + \alpha_2 a_t, \tag{21}$$

$$\alpha_{1,t} = \left(\alpha_1 + \frac{\alpha_3}{\alpha_4 \exp(-\alpha_5 m_t)} \right). \tag{22}$$

The approximating law of motion is again equal to a linear process and, thus, misses that the true process has a time-varying autoregressive coefficient. The values of the coefficients are chosen to be all positive so that shocks are more persistent when m_t takes on higher values. Two different sets of parameter values are considered. In the first set, the exogenous driving process, a_t, is serially correlated and $\rho_1 = 0.95$. In the second parameter set, a_t is not serially correlated. All other parameter values are identical, except that the value of σ_ε is adjusted to ensure that the standard deviation of m_t is the same in both cases and—as in the first experiment—equal to 0.025. With the first parameter set, the autoregressive coefficient, $\alpha_{1,t}$, varies between 0.90 and 0.95 (in a sample of 50,000 observations), whereas with the second parameter set it varies between 0.85 and 0.95. Although the variation is less in the first parameter set, the time-varying aspect of the autoregressive coefficient turns out to be more important for the first parameter set. The reason is that the persistence in a_t creates more persistence in $\alpha_{1,t}$ as well. The parameter values are reported in Table 2.

4.2.2 Experiment 2: Traditional accuracy test outcomes

Results for the traditional accuracy tests are reported in Table 3. The R^2 for the level equations are very high for both sets of parameter values. For $T = 3,000$, the minimum R^2 across Monte Carlo replications is above 0.9997 for both parameter sets. The R^2 for the first-difference regression gives different results. For Experiment 2.2, the R^2s are still high, with a minimum of 0.9975. For Experiment 2.1, however, the minimum is equal to 0.9385 and the average is equal to 0.977. Recall that the level and the first-difference regression are identical in all things that matter, including its predictions for m_{t+1}. The average standard error of the regression equation is equal to 0.021% and 0.03% for experiments 2.1 and 2.2 respectively. Low values that are consistent with the high R^2 values.

4.2.3 Experiment 2: New accuracy procedure

New accuracy procedure for Experiment 2: I - test outcomes. For $T = 3,000$, the average (median) value across Monte Carlo replications of $\widehat{\widehat{u}}^{\max}$ is equal to 1.86% (1.72%) and 1.83% (1.73%) for Experiment 2.1 and 2.2 respectively. These numbers are clearly not small and relative to the standard deviation of m_t, which is equal to 2.5%, they are huge.

Comparing the results for Experiment 2 with those of Experiment 1, two differences emerge. First, Experiment 2 makes much more clear than Experiment 1 that $\widehat{\widehat{u}}^{\max}$ is a much more powerful statistic than $\widehat{\widehat{u}}^{\text{ave}}$. Second, Experiment 2 makes clear the importance of using a long enough sample and/or do the accuracy test several times. That is, the minimum values of $\widehat{\widehat{u}}^{\max}$ make clear that the outcome of the accuracy test is reasonable in some Monte Carlo replications. In Experiment 2.1, the minimum value of $\widehat{\widehat{u}}^{\max}$ across Monte Carlo replications is equal to 0.46% when $T = 3,000$. This is considerably below the median value of 1.72%. The minimum value is equal to 2.01% when $T = 50,000$.

Table 5 reports the results for the statistics related to the 100-quarter ahead forecasts. Despite the differences in constructing the accuracy test, the results are again very similar to the results using $\widehat{\widehat{u}}^{\max}$ *as long as* one considers maximum forecast errors. The correlation coefficient for the 100-period ahead forecast and its realization is on average above 0.995

and, thus, clearly misses the inaccuracies of the approximating law of motion.

New accuracy procedure for Experiment 2: II - fundamental accuracy plot. Figure 4 plots the time series for m_{t+1} together with the series implied by the approximating law of motion when the approximating law is not updated by using the true observations for m_t as the explanatory variable. Panel A reports the results for Experiment 2.1 and Panel B for Experiment 2.2. The first 250 observations of the first Monte Carlo replication are used. The R^2 in this first Monte Carlo replication is equal to 0.9999 for Experiment 2.1 and 0.9998 for Experiment 2.2.

The results differ quite a bit across the two experiments. In Experiment 2.1, there is a systematic difference between the series generated by the true and the approximating law of motion. In Experiment 2.2, the series generated by the approximating law of motion follow the true series quite closely although there are some visible gaps. The figures seem to suggest that the results are much more accurate for Experiment 2.2 than for Experiment 2.1, whereas the new accuracy statistics indicate similar inaccuracies. Of course, the figure only plots an arbitrary part of an arbitrary Monte Carlo replication. Figure 5 plots for Experiment 2.2 the set of realizations for m_{t+1}, together with the corresponding values for $\widehat{\widehat{m}}_{t+1}$, for the Monte Carlo sample in which the largest value for this maximum error is obtained. The approximating law of motion of Experiment 2.2 can indeed do quite poorly for quite a long time period. Interestingly, before the true and the approximating law of motion diverge they track each other quite closely and the same is true afterwards. The approximating and the true law start to diverge when m_{t+1} takes on extremely low values. In particular, during severe downturns the approximating law of motion predicts much lower aggregate capital stocks.

New accuracy procedure for Experiment 2: III - evaluation. Figure 6 plots the impulse response functions using three different sets of initial conditions: the steady state value of m_t, two standard deviations above, and two standard deviations below m_t. The approximating law of motion is linear and initial conditions, thus, do not matter. Panel A reports the results for Experiment 2.1 and panel B reports the results for Experiment

2.2. The results for the approximating law of motion are those from the first Monte Carlo experiment.

For Experiment 2.2, the approximating law of motion does a good job in capturing the short-term response. It does a poor job, however, in capturing the speed at which the series revert to its pre-shock value; it overestimates the speed when the shock occurs at higher values and underestimates it at lower values. For Experiment 2.1, the approximating law of motion also does a good job in capturing the short-term responses, but after roughly five periods the approximating law of motion can generate quite different responses. The response generated by the approximating law of motion corresponds quite closely to the average response of the true law of motion, but this response is not the response observed when the shock occurs when the system is at the steady state.

5 Misleading answers from the R^2 test in actual application

In the last section, the specifications of the true laws of motion were chosen to highlight the shortcomings of the R^2 and $\widehat{\sigma}_u$ as accuracy measures. Although the processes chosen were regular processes, they did not come out of an actual economic model with heterogeneous agents. In the introduction, I did use data generated by a numerical solution to the individual policy rules from an actual economic model, but there was no feedback between aggregate and individual policy rules.

In this section, I present an example where relying on the R^2 in solving the model of Krusell and Smith (1998) leads to an inaccurate solution for a particular algorithm.[16] In particular, the R^2s of the approximating law of motion are above 0.99997. Nevertheless, there are nontrivial errors in the generated values of first-order moments.[17] The proposed numerical solution does fail, however, the new accuracy tests proposed in this paper.

[16] See den Haan, Judd, and Juillard (2009) for an exact description of the model.

[17] The inaccuracy seems to be present only in the mean values. There is no reason to believe, however, that inaccuracies will always only show up in first-order moments.

Algorithm and results. The example is from den Haan and Rendahl (2009), who show that the aggregate law of motion can be obtained by explicit aggregation of the individual policy rule. For explicit aggregation to work the individual policy rule has to be linear in the coefficients of the monomials of the individual state variables. If this is not the case, then an auxiliary individual policy rule has to be used that does have this form. The auxiliary rule is only used to obtain the aggregate law of motion and not to describe individual behavior. Because of the borrowing constraint, the individual policy rule of the unemployed agent in the Krusell-Smith economy has a kink and it would be costly to represent it as a function with the required functional form. In the first algorithm considered in den Haan and Rendahl (2009), auxiliary policy rules for the unemployed and employed agent are specified that are linear in the individual capital levels. From these den Haan and Rendahl (2009) derive an aggregate law of motion for the capital stock of the unemployed (K_u) and the employed (K_e) agents. The R^2s are equal to 0.999973 and 0.999997 for K_u and K_e respectively. Nevertheless, the solution is clearly inaccurate; the values of $\widehat{\widehat{u}}^{\max}$ $\left(\widehat{\widehat{u}}^{\text{ave}}\right)$are 1.68% (1.07%) and 1.45% (0.98%) for the law of motion of K_u and K_e respectively. The problem is that, although the one-period ahead forecasts are small, the bias is systematic and accumulates over time.

den Haan and Rendahl (2009) also presents a numerical solution that corrects for the bias and attains much lower values for $\widehat{\widehat{u}}^{\max}$. The accurate and inaccurate solution generate very similar second-order properties, but quite different first-order statistics. Without the bias correction, the mean capital stock implied by the approximating law of motion for aggregate capital is only 0.02% above the steady state capital stock, whereas with the bias correction, the implied mean capital stock is 0.28% above the steady state value. That is, the two numerical solutions give different answers to the question how uncertainty and imperfect risk sharing affect capital accumulation.

Moreover, these different aggregate laws of motion have a substantial effect on the individual policy rules. When the correction is imposed, then the mean capital stock implied by the individual policy function is 0.30% above the steady state value, that is, the number generated with the individual policy function is virtually the same as the

one generated with the aggregate law of motion. When the correction is not imposed, however, then the mean capital stock implied by the simulated individual choices is 1% above the steady state value.[18] Thus, although the aggregate law of motion obtained with the algorithm that does not apply the correction procedure has a very high R^2, the numerical solution gives very misleading answers regarding the question how the presence of idiosyncratic and aggregate uncertainty affects the mean capital stock in this model with incomplete markets.

6 Conclusion

Starting in the early nineties, numerous new numerical procedures have been developed to solve DSGE models. In those days, accuracy checks played a crucial role.[19] Even though models have become much more complex, researchers have become much less concerned about the accuracy of their numerical solutions. Krusell and Smith (1996, 1998) deserve credit for using a battery of accuracy tests when they used a new algorithm to solve a problem that had not been solved before. Unfortunately, this paper has documented that the two accuracy measures that were given most attention by Krusell and Smith (1996, 1998)—and that are now typically the only accuracy measures considered—are very weak tests. Now that it has been documented that high (low) values for the R^2 ($\widehat{\sigma}_u$) are pretty much meaningless, one should seriously question the accuracy of results presented when researchers only provide the values for the R^2 and $\widehat{\sigma}_u$.

This paper has also proposed a new accuracy procedure. The first part of the procedure consists of more powerful tests. More important than producing a particular number for an accuracy test, however, is an evaluation of the properties of the approximating aggregate law of motion and the aggregate law of motion implied by the individual policy rules.

[18] The difference in the mean implied by the aggregate and the individual laws of motion is, thus, roughly 0.98%, which is consistent with the values reported for $\widehat{\widehat{u}}^{\text{ave}}$ above.

[19] For example, accuracy tests played an important role in the comparison of numerical methods by Taylor and Uhlig (1991).

A Detailed results of the Monte Carlo Experiments

A.1 Experiment 1

Table 6 reports the differences between moments generated by the true and the approximating law of motion. The table consists of two parts. In the top panel, simulated series of m_t and $\widehat{\widehat{m}}_t$ are used to directly construct the moments. In the bottom panel, the unconditional moments implied by the true law of motion are compared with the unconditional moments implied by the estimated approximating law of motion.

One important observation is that for this exercise the sample size does matter. That is, even at a relatively large sample like 3,000 observations, the sampling variation has clearly not yet averaged out. This highlights that sampling variation disappears at a slow rate and using simulation procedures in numerical procedures may require the use of very large samples.

From a numerical point of view, some of the deviations are disturbing. For $T = 3,000$ the difference between the sample mean of the true m_t and the sample mean generated by the approximating law of motion is—averaged across Monte Carlo replications—equal to 0.17% for Experiment 1.2. This is not a good result, especially not if one realizes that this is 6.8% of the standard deviation of m_t. The maximum across Monte Carlo replications is equal to 0.71%, which is equal to 28.4% of the standard deviation of m_t. This maximum error for the mean is only reduced to 0.13% (or 5.2% of the standard deviation) when T is equal to $50,000$. Whether such differences in the mean are important depends on the context and how the approximating law of motion is used. A systematic error in the aggregate law of motion may lead to a systematic error in the individual policy rules, which in turn could increase the error in the aggregate law of motion.

Next, I check the difference between the true and the approximating law of motion for calculating impulse response functions. Table 7 reports the difference between the j^{th}-period impulse response according to the true and the approximating law of motion scaled by the true first-period response. The first two rows indicate that the approximating law of motion captures the first-period response quite well and that from an economic point

of view the differences are likely not to be important. The subsequent rows make clear that the approximating law of motion does a poor job in getting the shape of the impulse response function right. This should not be surprising, given that the hump that is present in the true impulse response function is impossible to capture with the approximating law of motion, which is a first-order autoregressive process. The point being made is, of course, not that some second-order processes cannot be accurately described with a first-order process. That is well known. The point is that the R^2 does not indicate that the first-order approximation is inaccurate.

A.2 Experiment 2

Table 8 documents the differences between the true and the approximating law of motion for the corresponding moments. Results do again vary by sample size, but the assessment about the accuracy of the approximating law of motion is the same for every sample size. Therefore, only the results for $T = 3,000$ are reported. The biggest differences are found for the standard deviation. Differences implied by the true and the approximating law of motion can be as high as 12.8% (6.97%) for Experiment 2.1 (2.2).

Table 9 reports summary statistics for the differences in the impulse response functions. The errors are much larger for Experiment 2.1. This is consistent with the "fundamental accuracy plot". One striking observation is that even for long-term responses there are large differences between the true and the approximating law of motion. In particular, for Experiment 2.1 the average (maximum) error for the 50-period response (as a fraction of the first-period response) is equal to 40.7% (85.2%).

The statistics in Table 9 are based on shocks that occur at the steady state. The true process is non-linear and the true impulses, thus, depend on the values of the state variables when the shock occurs. In particular, they are not that different if shocks occur at higher values, but they are quite different at lower values. As documented by the example in the main text, the differences between the true and the approximating law of motion are much larger when shocks occur at low values of m_t.

B Relation to Santos and Peralta-Alva (2005)

Some motivation for using the maximum prediction error in the accuracy test is found in Santos and Peralta-Alva (2005). Let s_t be equal to the vector $[m_t, a_t]$ and suppose that the true law of motion can be written as $m_{t+1} = \phi(s_t)$ with $s_t \in S$. Santos and Peralta-Alva (2005) show that one can use $\widehat{\widehat{u}}^{\max}$ to construct an error bound bound on moments of s_t as implied by the approximating law of motion $\bar{\phi}$.[20] Assume that there exist a constant γ such that

$$\mathrm{E}||\Phi(s,\varepsilon) - \Phi(s',\varepsilon)|| \leq \gamma||s - s'|| \text{ for all pairs } s, s', \tag{23}$$

where

$$\Phi(m, a, \varepsilon) = \begin{bmatrix} \phi(m,a) \\ \rho a + \varepsilon \end{bmatrix}. \tag{24}$$

Define d as follows

$$d = \max_{s,s'\varepsilon S} ||s - s'||. \tag{25}$$

Now consider the moments, $\mathrm{E}[f(s)]$, where $f(s)$ is a Lipschitz function with constant L.[21] Santos and Peralta-Alva (2005) show that one can bound the difference between the true moment, $\mathrm{E}[f(m_t, a_t)]$, and the expected value of the sample mean of $f(\widehat{\widehat{m}}_t, a_t)$ using the following inequality

$$\left| \mathrm{E}\left[f(m_t, a_t)\right] - \mathrm{E}\left[\frac{\sum_{t=\bar{T}}^{T} f(\widehat{\widehat{m}}_t, a_t)}{T - \bar{T}}\right] \right| \leq \frac{Ld\gamma^{T-\bar{T}}}{1-\gamma} + \frac{L\widehat{\widehat{u}}^{\max}}{1-\gamma}. \tag{26}$$

The first term on the right-hand side bounds the difference between the sample mean of $f(m_t, a_t)$ and $\mathrm{E}[f(m_t, a_t)]$. This term can be made arbitrarily small by letting T be large enough. The second term bounds the difference between the sample mean of $f(m_t, a_t)$ and the sample mean of $f(\widehat{\widehat{m}}_t, a_t)$. That is, it bounds the difference between the moment implied by the true law of motion and the moment implied by the approximating law of motion.

[20] The analysis allows for the possibility that ϕ depends on an additional explanatory variable such as m_{t-1}. One would have to extend the state space with the additional lag and $\bar{\phi}$ would restrict the coefficient on the additional term equal to zero.

[21] A real valued function f on S is called Lipschitz with constant L if $|f(s) - f(s')| \leq L||s - s'||$ for all pairs s and s' in S.

The formula of Santos and Peralta-Alva (2005) makes clear how $\widehat{\widehat{u}}^{\max}$ is a useful ingredient to bound moments of m_t. It also makes clear that a smaller value for $\widehat{\widehat{u}}^{\max}$ is needed when γ is closer to one. The higher the value of γ, the lower the value of $\widehat{\widehat{u}}^{\max}$ one needs to bound the error on the moment $\mathrm{E}[f(m_t, a_t)]$. Since aggregate series are often persistent, the value of γ could be quite high and to keep the upper bound low, one would need a very low value of $\widehat{\widehat{u}}^{\max}$. For example, suppose that one is interested in the mean so that $f(m_t) = m_t$ and $L = 1$. A value of $\widehat{\widehat{u}}^{\max} = 0.01\%$ and $\gamma = 0.99$, then imply a bound on the mean with a much larger value than $\widehat{\widehat{u}}^{\max}$ namely 1% (assuming that T is large enough to drive the first term to zero).

To use these bounds one needs an estimate for γ and L. To obtain an estimate for γ one can calculate the eigen values of the approximating law of motion and the law of motion for a_t and use the maximum. The value of L is equal to 1 when considering the mean. For other moments, one can obtain an estimate for L by calculating the maximum derivative of $f(s_t)$ across different values of s_t in S. Obviously, there are many unconditional moments one could look at. Besides standard unconditional moments of m_t, one can also consider impulse response functions.

This seems like a very useful framework to determine what reasonable values for $\widehat{\widehat{u}}^{\max}$ are. One first determines what the key moments of the true law of motion are, that one wants to capture. Next, one works backwards to determine what the value of $\widehat{\widehat{u}}^{\max}$ should be. For this to work in practice, one needs the bounds to be tight. Unfortunately, I found the bounds not to be tight in the examples used.

Consider Experiment 1.1. For the true driving process, the largest eigenvalue is equal to 0.9777.[22] The average value of $\widehat{\widehat{u}}^{\max}$ is equal to 0.83%. Using $\gamma = 0.9777$ implies an upper bound on the error of the implied mean of 37.2%.[23] Table 6 documents, however, that actual errors on the mean are way below this upper bound.

[22] In practice, one would have to use the approximating law of motion to obtain an estimate of the eigenvalue. Estimates across Monte Carlo replications vary from 0.9813 to 0.9823, so except for a small upward bias, the approximating law of motion is capable of capturing the dominant eigenvalue reasonably well.

[23] I ignore the first term in Equation (26), which captures sampling variation.

References

Algan, Y., O. Allais, and W. J. den Haan (2008): "Solving Heterogeneous-Agent Models with Parameterized Cross-Sectional Distributions," *Journal of Economic Dynamics and Control*, 32, 875–908.

——— (2009): "Solving the Incomplete Markets Model with Aggregate Uncertainty Using Parameterized Cross-Sectional Distributions," *Journal of Economic Dynamics and Control*, this issue.

den Haan, W. J. (1996): "Heterogeneity, Aggregate Uncertainty and the Short-Term Interest Rate," *Journal of Business and Economic Statistics*, 14, 399–411.

——— (1997): "Solving Dynamics Models with Aggregate Shocks and Heterogeneous Agents," *Macroeconomic Dynamics*, 1, 355–386.

den Haan, W. J., K. L. Judd, and M. Juillard (2009): "Computational Suite of Models with Heterogeneous Agents: Incomplete Markets and Aggregate Uncertainty," *Journal of Economic Dynamics and Control*, this issue.

den Haan, W. J., and P. Rendahl (2009): "Solving the Incomplete Markets Model with Aggregate Uncertainty Using Explicit Aggregation," *Journal of Economic Dynamics and Control*, this issue.

Gomes, F., and A. Michaelides (2008): "Asset Pricing with Limited Risk Sharing and Heterogeneous Agents," *Review of Financial Studies*, 21, 415–449.

Krueger, D., and F. Kubler (2004): "Computing Equilibrium in OLG Models with Stochastic Production," *Journal of Economic Dynamics and Control*, 28, 1411–1436.

Krusell, P., and A. A. Smith, Jr. (1996): "Income and Wealth Heterogeneity in the Macroeconomy," working paper version.

——— (1997): "Income and Wealth Heterogeneity, Portfolio Choice, and Equilibrium Asset Returns," *Macroeconomic Dynamics*, 1, 387–422.

——— (1998): "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, 106, 867–896.

——— (2006): "Quantitative Macroeconomic Models with Heterogeneous Agents," in *Advances in Economics and Econometrics: Theory and Applications, Ninth World Congress, Econometric Society Monographs*, ed. by R. Blundell, W. Newey, and T. Persson, pp. 298–340. Cambridge University Press.

Nakajima, M. (2007): "Business Cycles in the Equilibrium Model of Labor Market Search and Self-Insurance," unpublished manuscript, University of Illinois.

Reiter, M. (2001): "Estimating the Accuracy of Numerical Solutions to Dynamic Optimization Problems," unpublished manuscript, Universitat Pompeu Fabra.

——— (2002): "Recursive Computation of Heterogeneous Agent Models," mimeo.

——— (2009): "Solving Heterogeneous-Agent Models by Projection and Perturbation," *Journal of Economic Dynamics and Control*, forthcoming.

Ríos-Rull, J. V. (1997): "Computation of Equilibria in Heterogeneous Agent Models," *Federal Reserve Bank of Minneapolis Staff Report*, 231, 238–264.

Santos, M. (2000): "Accuracy of Numerical Solutions Using the Euler Equations Residual," *Econometrica*, 68, 1377–1402.

Santos, M. S., and A. Peralta-Alva (2005): "Accuracy of Simulations for Stochastic Dynamic Models," *Econometrica*, 73, 1939–1976.

Silos, P. (2007): "Housing, Portfolio Choice and the Macroeconomy," *Journal of Economic Dynamics and Control*, 31, 2774–2801.

Storesletten, K., C. I. Telmer, and A. Yaron (2007): "Asset Pricing with Idiosyncratic Risk and Overlapping Generations," *Review of Economic Dynamics*, 10, 519–548.

TAYLOR, J. B., AND H. UHLIG (1991): "Solving Nonlinear Stochastic Growth Models: A Comparison of Alternative Solution Methods," *Journal of Business and Economic Statistics*, 8, 1–17.

VERBEEK, M. (2000): *A Guide to Modern Econometrics*. John Wiley and Sons, Chichester.

YOUNG, E. R. (2005): "Approximate Aggregation," unpublished manuscript, University of Virginia.

——— (2009): "Solving the Incomplete Markets Model with Aggregate Uncertainty Using the Krussell-Smith Algorithm and Non-Stochastic Simulations," *Journal of Economic Dynamics and Control*, this issue.

Table 1: Meaninglessness of the R^2

			implied properties	
equation	R^2	$\widehat{\sigma}_u$	mean	stand. dev.
$\alpha_3 = 0.96404$ (fitted regression)	0.99999729	4.1×10^{-5}	3.6723	0.0248
$\alpha_3 = 0.954187$	0.99990000	2.5×10^{-4}	3.6723	0.0217
$\alpha_3 = 0.9324788$	0.99900000	7.9×10^{-4}	3.6723	0.0174
$\alpha_3 = 0.8640985$	0.99000000	2.5×10^{-3}	3.6723	0.0113

Notes: The first row corresponds to the fitted regression equation. The subsequent rows are based on aggregate laws of motion in which the value of α_3 is changed until the indicated level of the R^2 is obtained; α_1 is adjusted to keep the fitted mean capital stock equal.

Table 2: Parameter Values

Parameter	Experiment 1		Experiment 2	
	1.1	1.2	2.1	2.2
α_0	0	0		0
α_1	1.08	1.38	0.65	0.65
α_2	1	1	1	1
$\alpha 3$	-0.1	-0.4	0.3	0.3
α_4	-	-	0.01	0.01
α_5	-	-	50	50
ρ_0	0	0	0	0
ρ_1	0	0	0.95	0
σ	0.00472	0.15436	$6.3891 * 10^{-4}$	$8.616 * 10^{-3}$

Notes: All parameter sets imply a standard deviation for the underlying series equal to 2.5%.

Table 3: Traditional Accuracy Tests

	Experiment 1.1		Experiment 1.2	
T	$3,000$	$50,000$	$3,000$	$50,000$
average R^2 (level)	0.9996	0.9996	0.9952	0.9955
minimum R^2 (level)	0.9995	0.9996	0.9940	0.9951
average R^2 (Δ)	0.9901	0.9901	0.8413	0.8411
minimum R^2 (Δ)	0.9901	0.9901	0.8408	0.8410
average $\widehat{\sigma}_u$	0.047%	0.047%	0.168%	0.168%
maximum $\widehat{\sigma}_u$	0.049%	0.048%	0.174%	0.170%
	Experiment 2.1		Experiment 2.2	
T	$3,000$	$50,000$	$3,000$	$50,000$
average R^2 (level)	0.99993	0.99993	0.99986	0.99986
minimum R^2 (level)	0.99983	0.99991	0.99971	0.99982
average R^2 (Δ)	0.97695	0.97559	0.99879	0.99880
minimum R^2 (Δ)	0.93847	0.96828	0.99750	0.99847
average $\widehat{\sigma}_u$	0.021%	0.022%	0.030%	0.030%
maximum $\widehat{\sigma}_u$	0.034%	0.025%	0.044%	0.034%

Notes: The standard deviation of the true series is equal to 2.5%. R^2 for the level (Δ) regression is based on a regression with m_{t+1} ($m_{t+1} - m_t$) as the dependent variable.

Table 4: New Accuracy Tests

	Experiment 1.1		Experiment 1.2	
T	$3,000$	$50,000$	$3,000$	$50,000$
average $\widehat{\widehat{u}}^{\max}$	0.83%	1.03%	3.34%	4.10%
median $\widehat{\widehat{u}}^{\max}$	0.82%	1.01%	3.28%	4.04%
minimum $\widehat{\widehat{u}}^{\max}$	0.57%	0.85%	2.48%	3.40%
average $\widehat{\widehat{u}}^{\text{ave}}$	0.21%	0.20%	0.83%	0.80%
minimum $\widehat{\widehat{u}}^{\text{ave}}$	0.16%	0.19%	0.67%	0.76%
	Experiment 2.1		Experiment 2.2	
T	$3,000$	$50,000$	$3,000$	$50,000$
average $\widehat{\widehat{u}}^{\max}$	1.86%	3.20%	1.83%	3.07%
median $\widehat{\widehat{u}}^{\max}$	1.72%	3.21%	1.73%	2.95%
minimum $\widehat{\widehat{u}}^{\max}$	0.46%	2.01%	0.59%	2.16%
average $\widehat{\widehat{u}}^{\text{ave}}$	0.21%	0.20%	0.17%	0.17%
minimum $\widehat{\widehat{u}}^{\text{ave}}$	0.11%	0.18%	0.12%	0.15%

Notes: $\widehat{\widehat{u}}$ stands for the difference between m and $\widehat{\widehat{m}}$; the latter uses lagged values of $\widehat{\widehat{m}}$ and not lagged values of m as values for explanatory variable. The superscript indicates whether the maximum or the average is taken. The standard deviation of the true series is equal to 2.5%.

Table 5: Multi-Step Forecasts & Errors

	Experiment 1.1		Experiment 1.2	
T	$3,000$	$50,000$	$3,000$	$50,000$
average $\widehat{u}^{\tau,\max}$	0.81%	1.01%	3.35%	4.13%
median $\widehat{u}^{\tau,\max}$	0.81%	0.98%	3.33%	4.04%
minimum $\widehat{u}^{\tau,\max}$	0.65%	0.88%	2.76%	3.63%
average $\widehat{u}^{\tau,\text{ave}}$	0.17%	0.20%	0.70%	0.81%
minimum $\widehat{u}^{\tau,\text{ave}}$	0.16%	0.19%	0.63%	0.79%
average correlation$(m_{t+\tau}, \widehat{m}_{t+\tau,t})$	0.9948	0.9950	0.9101	0.9128
minimum correlation$(m_{t+\tau}, \widehat{m}_{t+\tau,t})$	0.9929	0.9946	0.8867	0.9071
	Experiment 2.1		Experiment 2.2	
T	$3,000$	$50,000$	$3,000$	$50,000$
average $\widehat{\widehat{u}}^{\max}$	1.79%	3.15%	1.79%	3.15%
median $\widehat{\widehat{u}}^{\max}$	1.65%	2.95%	1.65%	2.95%
minimum $\widehat{\widehat{u}}^{\max}$	0.62%	2.14%	0.62%	2.15%
average $\widehat{\widehat{u}}^{\text{ave}}$	0.14%	0.17%	0.14%	0.17%
minimum $\widehat{\widehat{u}}^{\text{ave}}$	0.08%	0.15%	0.08%	0.15%
average correlation$(m_{t+\tau}, \widehat{m}_{t+\tau,t})$	0.9953	0.9953	0.9953	0.9953
minimum correlation$(m_{t+\tau}, \widehat{m}_{t+\tau,t})$	0.9832	0.9942	0.9832	0.9942

Notes: This table compares τ-period ahead forecasts, $\widehat{m}_{t+\tau,t}$, with the realization, $m_{t+\tau}$, and reports properties of the forecast error. τ is equal to 100. The standard deviation of the true series is equal to 2.5%.

Table 6: Moments - Experiment 1

	1.1		1.2	
T	$3,000$	$50,000$	$3,000$	$50,000$
average error for $\widehat{\widehat{\mu}}_m$	0.05%	0.01%	0.17%	0.04%
maximum error for $\widehat{\widehat{\mu}}_m$	0.21%	0.04%	0.71%	0.13%
average % error for $\widehat{\widehat{\sigma}}_m$	1.40%	0.55%	9.68%	8.40%
maximum % error for $\widehat{\widehat{\sigma}}_m$	3.74%	1.51%	19.4%	12.1%
average % error for $\widehat{\widehat{\rho}}_{-1}$	0.05%	0.01%	0.15%	0.04%
maximum % error for $\widehat{\widehat{\rho}}_{-1}$	0.17%	0.04%	0.53%	0.14%
average % error for $\widehat{\widehat{\rho}}_{-2}$	0.37%	0.37%	1.10%	1.17%
maximum % error for $\widehat{\widehat{\rho}}_{-2}$	0.65%	0.45%	2.02%	1.44%
average % error for $\widehat{\widehat{\rho}}_{m,a}$	1.18%	0.52%	10.0%	9.13%
maximum % error for $\widehat{\widehat{\rho}}_{m,a}$	3.97%	1.30%	29.2%	12.58%
average error for $\widetilde{\widehat{\mu}}_m$	0.04%	0.01%	0.12%	0.03%
maximum error for $\widetilde{\widehat{\mu}}_m$	0.13%	0.03%	0.39%	0.10%
average % error for $\widetilde{\widehat{\sigma}}_m$	0.79%	0.49%	7.07%	8.32%
maximum % error for $\widetilde{\widehat{\sigma}}_m$	2.28%	1.14%	13.9%	10.4%
average % error for $\widetilde{\widehat{\rho}}_{-1}$	0.03%	0.01%	0.08%	0.02%
maximum % error for $\widetilde{\widehat{\rho}}_{-1}$	0.07%	0.02%	0.20%	0.06%
average % error for $\widetilde{\widehat{\rho}}_{-2}$	0.39%	0.37%	1.21%	1.15%
maximum % error for $\widetilde{\widehat{\rho}}_{-2}$	0.51%	0.40%	1.55%	1.24%
average % error for $\widetilde{\widehat{\rho}}_{m,a}$	0.79%	0.50%	7.95%	9.14%
maximum % error for $\widetilde{\widehat{\rho}}_{m,a}$	2.14%	1.16%	15.4%	11.37%

Notes: The symbol $\widehat{\ }$ indicates that simulated series are used to construct the statistics for the true and the approximating law of motion (using a sample length equal to the one used to estimate the parameters of the approximating law of motion). The symbol $\widetilde{\ }$ indicates that the true moments implied by the true and approximating laws of motion are used. The standard deviation of the underlying process is equal to 2.5%.

Table 7: Impulse Response Functions - Experiment 1

	1.1			1.2		
T	1000	$3,000$	$50,000$	1000	$3,000$	$50,000$
average error for $\widetilde{\widehat{i}}_1$	0.28%	0.16%	0.04%	1.24%	0.75%	0.17%
maximum error for $\widetilde{\widehat{i}}_1$	0.76%	0.56%	0.11%	3.69%	2.16%	0.48%
average error for $\widetilde{\widehat{i}}_2$	9.74%	9.76%	9.92%	39.22%	39.2%	39.4%
maximum error for $\widetilde{\widehat{i}}_2$	10.6%	10.2%	10.2%	43.0%	41.3%	39.8%
average error for $\widetilde{\widehat{i}}_3$	10.1%	10.2%	10.4%	53.0%	53.0%	53.2%
maximum error for $\widetilde{\widehat{i}}_3$	11.0%	10.7%	10.4%	56.8%	55.2%	53.7%
average error for $\widetilde{\widehat{i}}_4$	9.59%	9.65%	9.72%	56.3%	56.3%	56.6%
maximum error for $\widetilde{\widehat{i}}_4$	10.5%	10.2%	9.86%	60.0%	58.7%	57.0%
average error for $\widetilde{\widehat{i}}_5$	8.96%	9.04%	9.12%	55.3%	55.4%	55.7%
maximum error for $\widetilde{\widehat{i}}_5$	9.98%	9.62%	9.28%	59.0%	57.8%	56.2%
average error for $\widetilde{\widehat{i}}_{50}$	4.44%	3.96%	3.67%	19.5%	18.4%	17.5%
maximum error for $\widetilde{\widehat{i}}_{50}$	7.85%	5.28%	4.01%	29.6%	22.9%	18.7%
average error for $\widetilde{\widehat{i}}_{100}$	4.90%	4.49%	4.26%	20.7%	19.3%	18.4%
maximum error for $\widetilde{\widehat{i}}_{100}$	7.91%	5.54%	4.53%	31.5%	23.7%	19.6%

Notes: Here i_j indicates the j^{th}-period response. Errors are calculated as a fraction of the true first-period response. The standard deviation of the underlying process is equal to 2.5%.

Table 8: Moments - Experiment 2

	2.1	2.2
T	$3,000$	$3,000$
average error for $\widehat{\widehat{\mu}}_m$	0.05%	0.04%
maximum error for $\widehat{\widehat{\mu}}_m$	0.17%	0.12%
average % error for $\widehat{\widehat{\sigma}}_m$	3.03%	1.61%
maximum % error for $\widehat{\widehat{\sigma}}_m$	12.8%	6.97%
average % error for $\widehat{\widehat{\rho}}_{-1}$	0.01%	0.21%
maximum % error for $\widehat{\widehat{\rho}}_{-1}$	0.09%	0.85%
average % error for $\widehat{\widehat{\rho}}_{-2}$	0.03%	0.41%
maximum % error for $\widehat{\widehat{\rho}}_{-2}$	0.15%	1.68%
average % error for $\widehat{\widehat{\rho}}_{m,a}$	0.70%	1.51%
maximum % error for $\widehat{\widehat{\rho}}_{m,a}$	2.08%	4.31%
average error for $\widetilde{\widehat{\mu}}_m$	0.23%	0.25%
maximum error for $\widetilde{\widehat{\mu}}_m$	0.34%	0.32%
average % error for $\widetilde{\widehat{\sigma}}_m$	3.98%	1.97%
maximum % error for $\widetilde{\widehat{\sigma}}_m$	9.87%	5.20%
average % error for $\widetilde{\widehat{\rho}}_{-1}$	0.02%	0.27%
maximum % error for $\widetilde{\widehat{\rho}}_{-1}$	0.03%	0.66%
average % error for $\widetilde{\widehat{\rho}}_{-2}$	0.06%	0.55%
maximum % error for $\widetilde{\widehat{\rho}}_{-2}$	0.09%	1.29%
average % error for $\widetilde{\widehat{\rho}}_{m,a}$	1.53%	1.13%
maximum % error for $\widetilde{\widehat{\rho}}_{m,a}$	3.24%	3.62%

Notes: The symbol $\widehat{\ }$ indicates that simulated series are used to construct the statistics for the true and the approximating law of motion using a sample length equal to the one used to estimate the parameters of the approximating law of motion. The symbol $\widetilde{\widehat{\ }}$ indicates that the true moments implied by the true and approximating laws of motion are used. The standard deviation of the underlying process is equal to 2.5%.

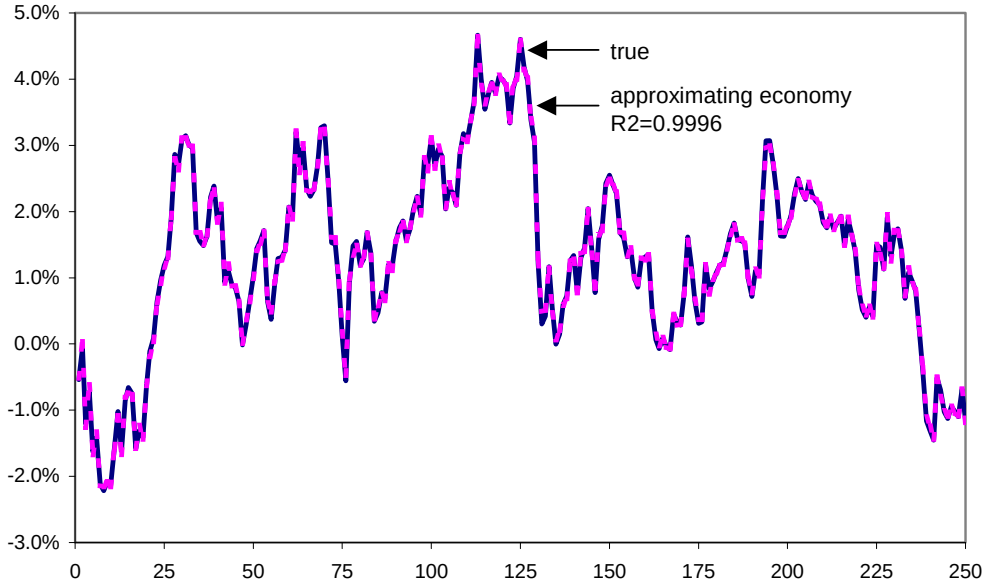
Table 9: Impulse Response Functions - Experiment 2

	2.1	2.2
T	$3,000$	$3,000$
average error for $\widetilde{\widehat{i}}_1$	2.22%	0.05%
maximum error for $\widetilde{\widehat{i}}_1$	6.19%	0.22%
average error for $\widetilde{\widehat{i}}_2$	4.54%	0.71%
maximum error for $\widetilde{\widehat{i}}_2$	12.5%	1.23%
average error for $\widetilde{\widehat{i}}_3$	6.93%	1.34%
maximum error for $\widetilde{\widehat{i}}_3$	18.7%	2.19%
average error for $\widetilde{\widehat{i}}_4$	9.37%	1.89%
maximum error for $\widetilde{\widehat{i}}_4$	24.9%	3.05%
average error for $\widetilde{\widehat{i}}_5$	11.8%	2.38%
maximum error for $\widetilde{\widehat{i}}_5$	31.0%	3.83%
average error for $\widetilde{\widehat{i}}_{50}$	40.7%	2.01%
maximum error for $\widetilde{\widehat{i}}_{50}$	85.2%	3.08%
average error for $\widetilde{\widehat{i}}_{100}$	9.11%	0.22%
maximum error for $\widetilde{\widehat{i}}_{100}$	18.3%	0.32%

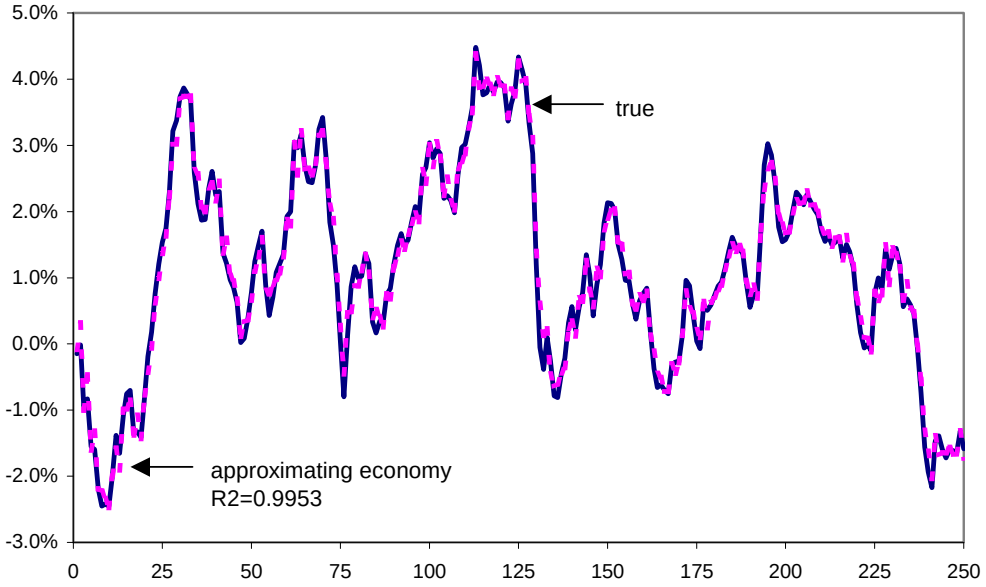
Notes: Here i_j indicates the j^{th}-period response. Errors are calculated as a fraction of the true first-period response. The standard deviation of the underlying process is equal to 2.5%.

Figure 1: True and predicted K' (updated values for K used in approximating law)

Panel A: Experiment 1.1



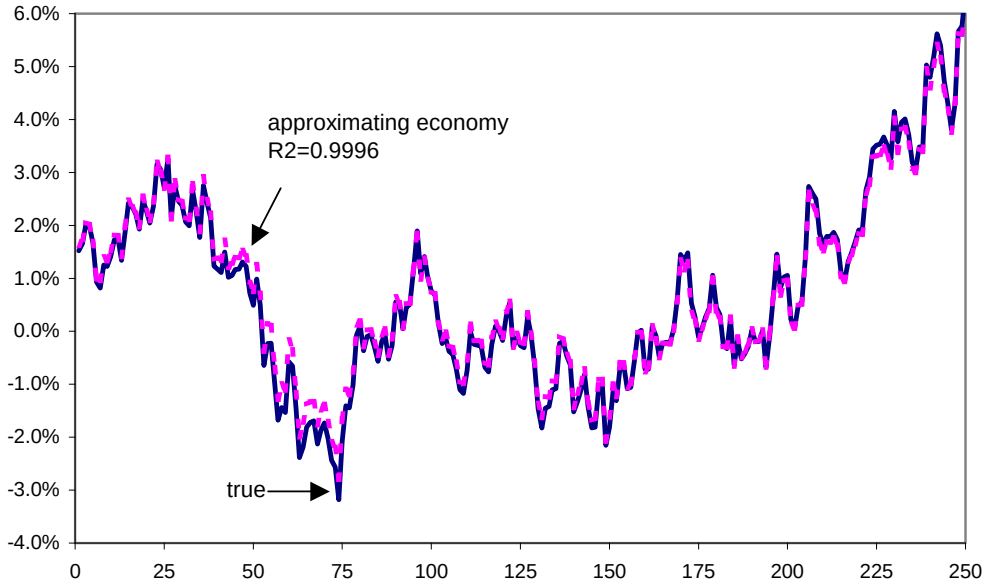
Panel B: Experiment 1.2



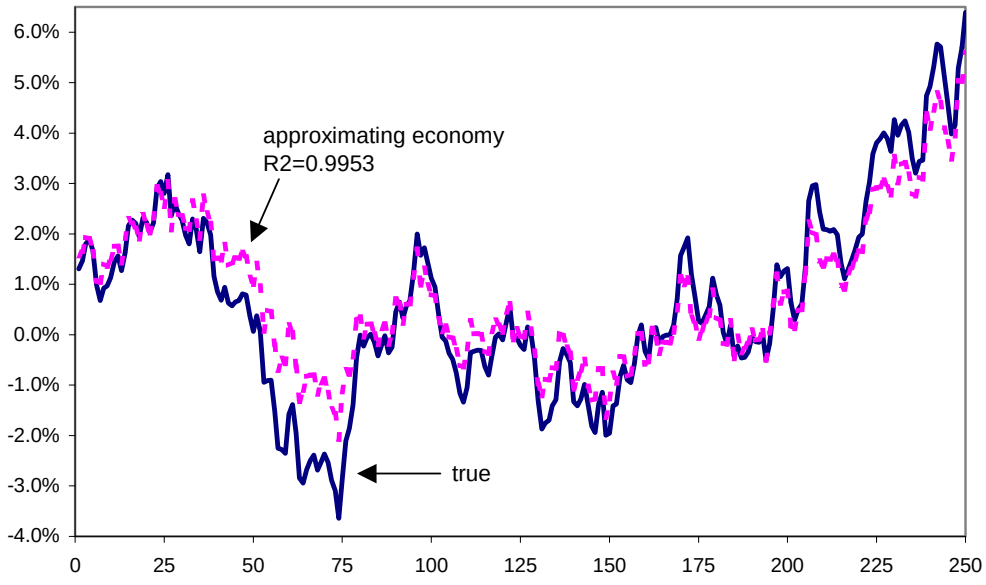
Notes: The graph plots the first 250 observations of the first Monte Carlo replication. The R^2 refers to the fit of the complete sample with 3,000 observations.

Figure 2: True and predicted *K'* (updated values for *K not* used in approximating law)

Panel A: Experiment 1.1



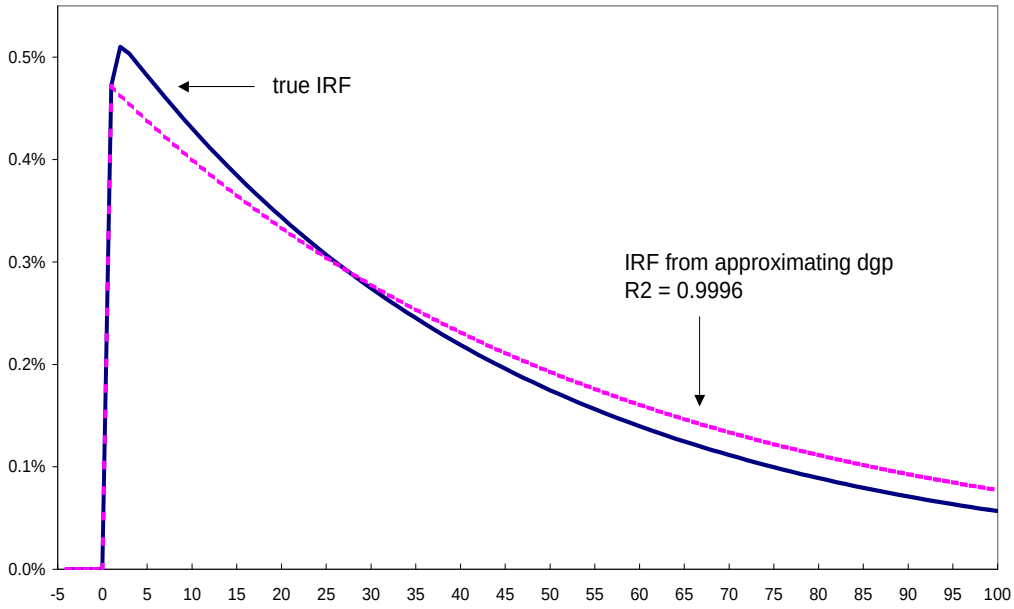
Panel B: Experiment 1.2



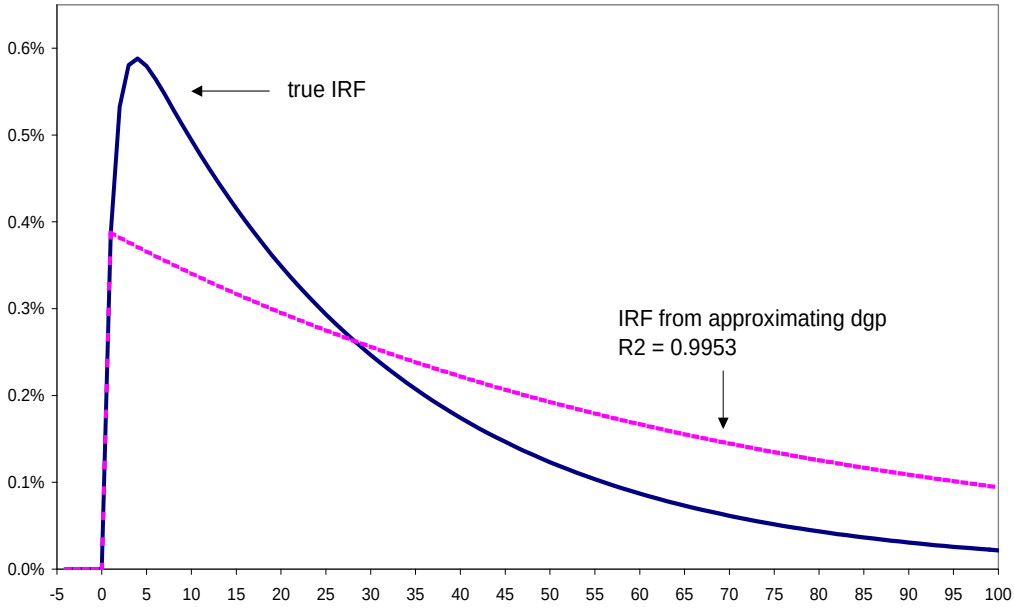
Notes: This is the fundamental accuracy plot. The graph plots the first 250 observations of the first Monte Carlo replication. The R^2 refers to the fit when the approximating law of motion is updated using the true observations of *K* as explanatory variables as in Figure 1.

Figure 3: Impulse response functions

Panel A: Experiment 1.1



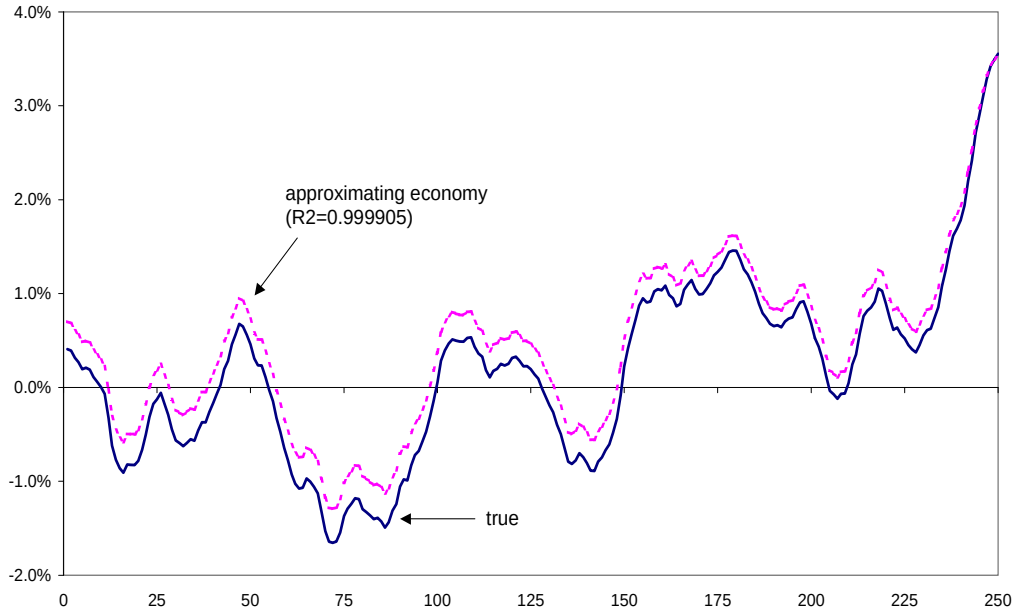
Panel B: Experiment 1.2



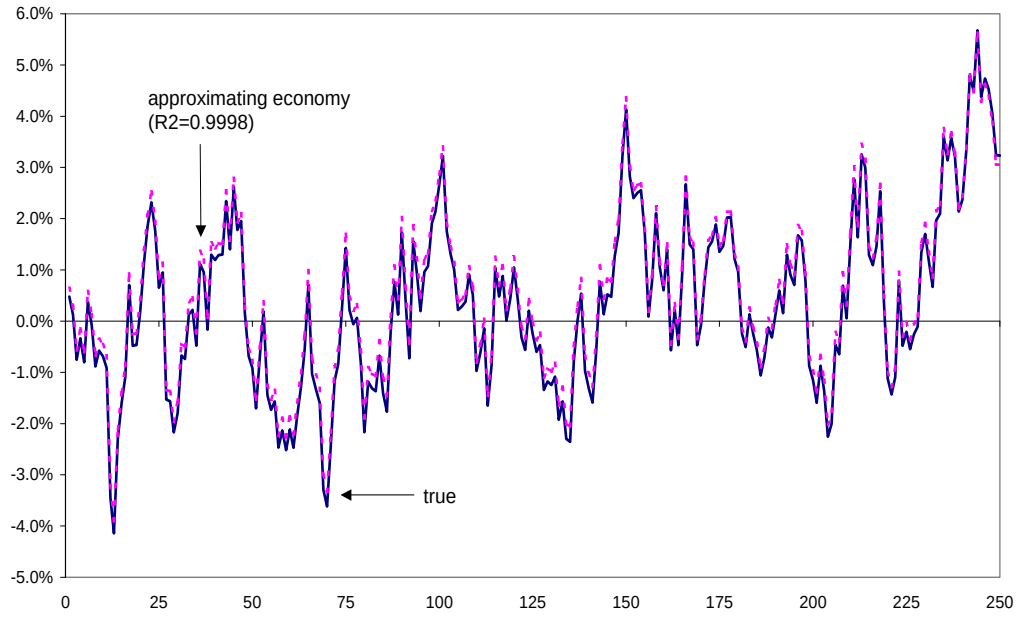
Notes: The graph plots the response of K_t in response to a productivity shock. The R^2 refers to the fit in simulated data when the approximating law is updated using the true observations of K as explanatory variable.

Figure 4: True and predicted *K'* (updated values for *K not* used in approximating law)

Panel A: Experiment 2.1



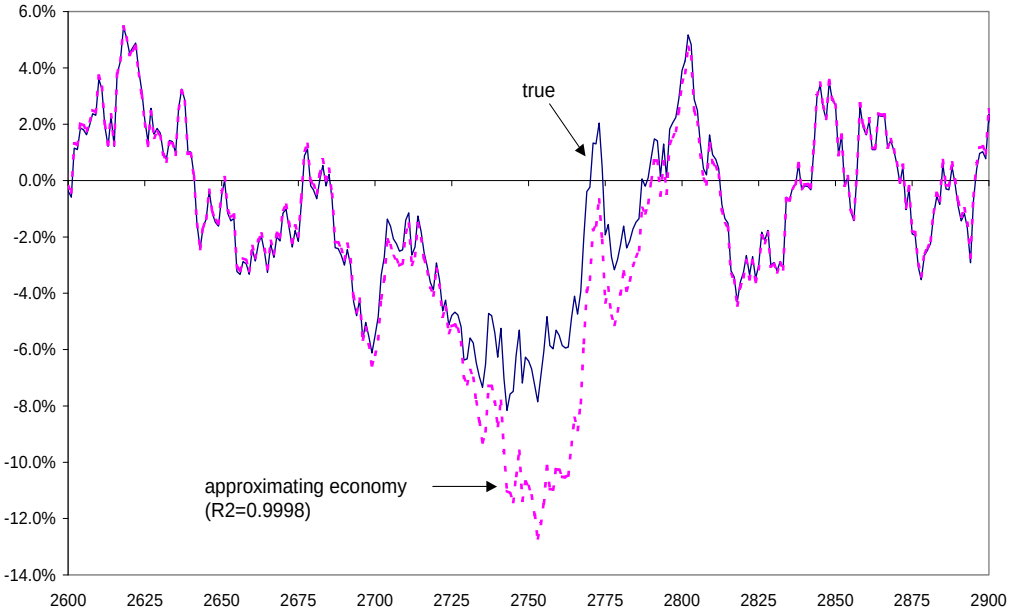
Panel B: Experiment 2.2



Notes: This is the fundamental accuracy plot. The graph plots the first 250 observations of the first Monte Carlo replication. The R^2 refers to the fit when the approximating law of motion is updated using the true observations of *K* as explanatory variables.

Figure 5: True and predicted *K* - (updated values for *K not* used in approximating law)

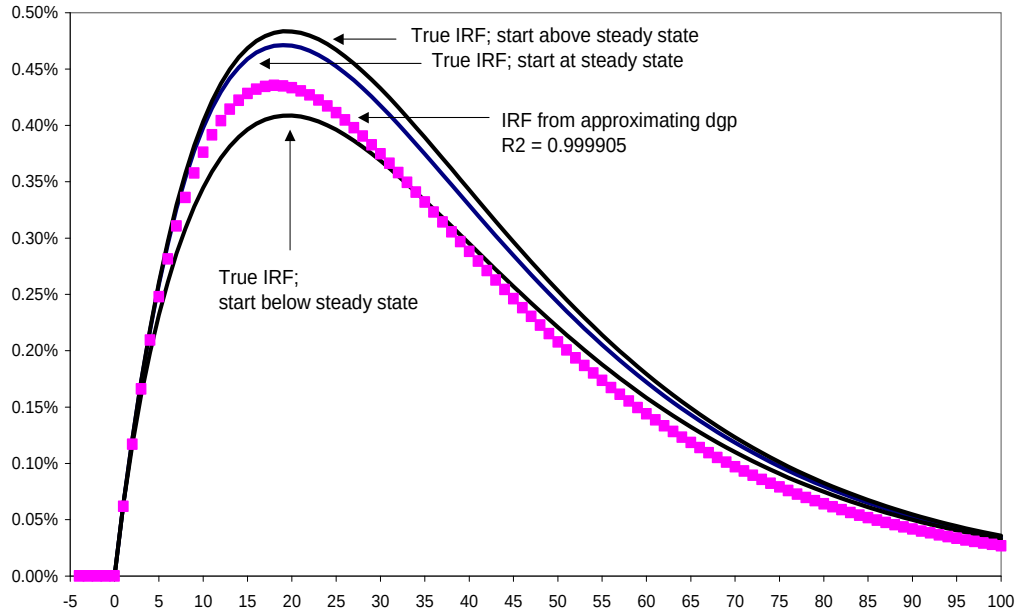
Experiment 2.2 – part of simulation where maximum error occurs



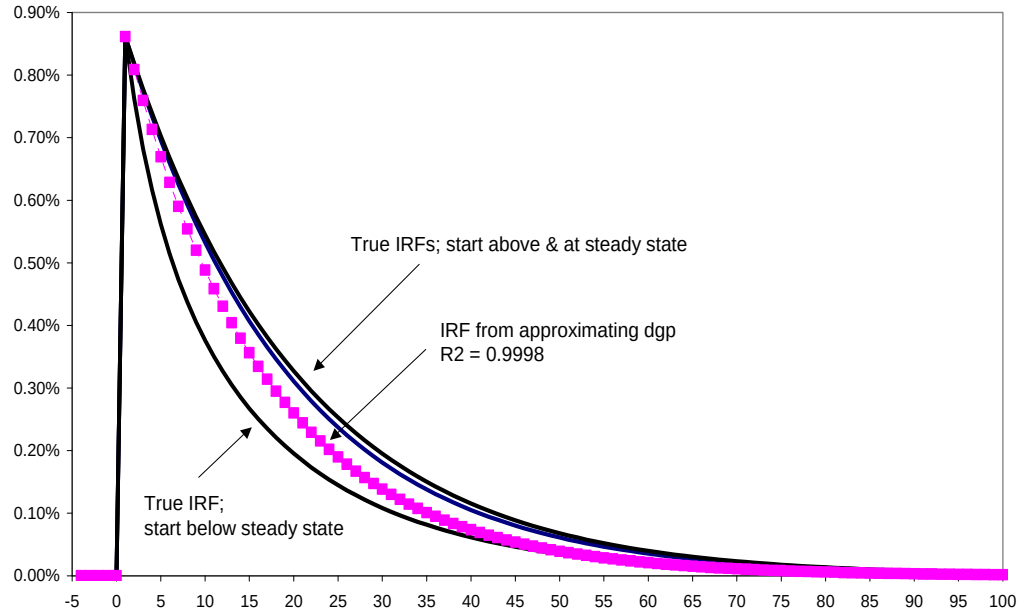
Notes: This is the fundamental accuracy plot. The R^2 refers to the fit when the approximating law of motion is updated using the true observations of *K* as explanatory variables.

Figure 6: Impulse response functions

Panel A: Experiment 2.1



Panel B: Experiment 2.2



Notes: The graph plots the response of K_t in response to a productivity shock. The true law of motion is non-linear and the IRFs depend on initial conditions; the IRFs of the approximating law of motion do not depend on initial conditions, because the approximating law of motion is linear. The R^2 refers to the fit in simulated data when the approximating law is updated using the true observations of K as explanatory variable.