Blasques, Francisco; Nientker, Marc

Working Paper
Transformed Perturbation Solutions for Dynamic Stochastic General Equilibrium Models

Tinbergen Institute Discussion Paper, No. TI 2019-012/III

Provided in Cooperation with:
Tinbergen Institute, Amsterdam and Rotterdam

TI 2019-012/III
Tinbergen Institute Discussion Paper

Transformed Perturbation Solutions for Dynamic Stochastic General Equilibrium Models

Francisco (F.) Blasques[1]
Marc Nientker[1]

[1] VU Amsterdam, The Netherlands

Transformed Perturbation Solutions for Dynamic Stochastic General Equilibrium Models*

Francisco Blasques[(a,b)] and Marc Nientker[(a)†]

(a) VU University Amsterdam
(b) Tinbergen Institute

December 13, 2018

Abstract

This paper introduces a new solution method for Dynamic Stochastic General Equilibrium (DSGE) models that produces non explosive paths. The proposed solution method is as fast as standard perturbation methods and can be easily implemented in existing software packages like *Dynare* as it is obtained directly as a transformation of existing perturbation solutions proposed by Judd and Guu (1997) and Schmitt-Grohe and Uribe (2004), among others. The transformed perturbation method shares the same advantageous function approximation properties as standard higher order perturbation methods and, in contrast to those methods, generates stable sample paths that are stationary, geometrically ergodic and absolutely regular. Additionally, moments are shown to be bounded. The method is an alternative to the pruning method as proposed in Kim et al. (2008). The advantages of our approach are that, unlike pruning, it does not need to sacrifice accuracy around the steady state by ignoring higher order effects and it delivers a policy function. Moreover, the newly proposed solution is always more accurate globally than standard perturbation methods. We demonstrate the superior accuracy of our method in a range of examples.

1 Introduction

Since the seminal paper of Kydland and Prescott (1982) many different methods have been proposed to approximate the solution of Dynamic Stochastic General Equilibrium (DSGE) models, see for example Taylor and Uhlig (1990), Christiano and Fisher (2000) and Aruoba et al. (2006) for comparison studies. It is well known that, in most cases, closed form analytical solutions do not exist, and hence we need numerical solution methods.

When selecting solution methods, two properties are of main interest: *speed and accuracy*. On the one hand, arbitrarily accurate solution algorithms such as value function iteration (Bertsekas, 1987) and projection methods (Judd, 1992) have existed for a long time. However, such methods need long computing times. This is problematic, especially when one is interested in estimating a DSGE model, since then the solution will have

*We thank Michel Juillard, Sergey Ivashchenko and the attendants of the "Advances in solution methods" session of the 14th Dynare Conference for helpful comments and suggestions. We thank Martin Andreasen and Wouter Den Haan for making code available that aided the paper.

†Corresponding author, email: m.h.c.nientker@vu.nl.

to be computed for a range of different parameter values. On the other hand, very fast solution methods such as linearization (Blanchard and Kahn, 1980) and higher-order perturbation methods (Judd and Guu, 1997; Schmitt-Grohé and Uribe, 2004) are available. These methods approximate the solution by taking a Taylor series expansion around the deterministic steady state. Unfortunately, despite being very fast, perturbation methods also have important limitations.

Linearization, or first order perturbation, can be very inaccurate and is often too simplistic from an economic perspective. For example, linear solutions are certainty equivalent and therefore miss potential volatility dynamics in the innovations. That means that one needs higher order perturbation methods for risk to matter, which affects a multitude of topics. For instance, this is a relevant limitation when attempting to model time varying risk premia as in Fernández-Villaverde et al. (2011); Rudebusch and Swanson (2012); Fernández-Villaverde et al. (2015) and requires a perturbation approximation of at least third order to be solved. Similarly, linearization is highly inaccurate when comparing welfare across different environments and can lead to paradoxical results (Tesar, 1995). Kim and Kim (2003b) show that a welfare comparison based on a linear approximation of the policy function may yield spurious results in a two-agent economy and that perturbation approximations of at least second order are required. Some welfare studies that use higher order perturbation approximations can be found in Kollmann (2002), Kim and Kim (2003a) and Bergin et al. (2007).[1] Finally, Van Binsbergen et al. (2012) discuss the need for higher order perturbation solutions to study consumer risk aversion.

The speed of perturbation methods and their ability to locally capture important nonlinear dynamics renders high-order perturbation a popular solution method. However, higher-order perturbation is an unattractive approximation method from a global perspective as it defines an unstable dynamic system which produces explosive paths. In fact, one can commonly show that sample paths generated using higher-order perturbations diverge to infinity almost surely, even if the true policy function implies stable dynamics with nonexplosive paths. This problem is outlined in Aruoba et al. (2006) and Den Haan and De Wind (2010) and encountered in Fahr and Smets (2010) and Den Haan and De Wind (2012), among others. See Section 3.3.2 and Section 5 in Den Haan and De Wind (2010) for extensively discussed examples.

In order to deal with the unstable dynamics of higher-order perturbation solutions, Kim et al. (2008) proposed the pruning method. The pruning method has been successfully implemented in software packages and effectively solves the problem of explosive dynamics; see also Andreasen et al. (2017) for recent results on the stability and stationarity of pruned solutions. However, pruned solutions must sacrifice local approximation accuracy for stability. Den Haan and De Wind (2010) show that pruning *"creates large systematic distortions"*. Furthermore, pruning is a simulation-based approximation and hence does not provide a policy function. In fact, approximations based on the pruning procedure contain different updates for identical values of the model's original state variables. This means that *"the implied policy rule is not even a function of the model's state variables"* (Den Haan and De Wind, 2010).

Our paper introduces a new transformed perturbation solution method for DSGE models that is designed to avoid explosive paths produced by higher-order perturbation solutions. Transformed perturbation is as fast as standard perturbation methods and can be easily implemented in existing software packages like *Dynare* as it is obtained directly

[1] Woodford (2002) discusses a set of assumptions that ensure first order approximations are sufficient.

as a transformation of existing perturbation solutions. The new method transforms the standard perturbation approximation by replacing higher order monomials in the Taylor expansion with transformed ones that are based on the transformed polynomials introduced in Blasques (2014). Transformed polynomial functions share the same fundamental approximation properties as polynomial functions. Blasques (2014) shows that transformed polynomials are dense in the space of continuous functions and attain the same rates of convergence as polynomials in Sobolev spaces of n times continuously differentiable functions. Additionally, in this paper, transformed perturbation is shown to converge on analytic function domains and to have the same excellent local properties as the standard perturbation method for continuously differentiable functions of appropriate order. From a global perspective however, transformed perturbation performs infinitely better than regular perturbation, because it provides a way of scaling down the higher order perturbation terms that cause explosive behavior when the solution path moves far away from the steady state. That way, transformed perturbation can be guaranteed to not create additional fixed points if the Blanchard-Kahn conditions are satisfied. Moreover, unlike pruning, the new solution method does not need to sacrifice accuracy by ignoring higher order effects. Additionally, transformed perturbation is guaranteed to be always more accurate than standard perturbation methods, which is not the case for pruned solutions. Finally, in contrast to pruning, transformed perturbation also has the advantage of delivering a policy function from which the simulations are drawn.

In this paper, we prove that transformed perturbation produces non explosive paths and that solutions are stable and strictly stationary ergodic with bounded moments. Additionally, we show that solution paths exhibit fading memory (i.e. geometric ergodicity and absolute regularity or β-mixing) and that sample moments of the process converge exponentially fast to the moments of the solution. These are crucial properties for conducting simulation-based estimation of parameters and simulation-based analysis of the DSGE model. Overall, this renders the transformed polynomial solution attractive from both a practical and theoretical stand-point.

We demonstrate the accuracy of the transformed perturbation method extensively for two nonlinear DSGE models in which higher order perturbation is infeasible. We compare second order transformed perturbation to first order perturbation and second order pruning. The first model is a partial equilibrium model in which agents face idiosyncratic income risk, introduced in Deaton Angus (1991) and Den Haan and De Wind (2012). For this model we find that sample path errors of our method are less than half of those of pruning and up to six times less than those for first order perturbation. This then results into sample moments of the transformed perturbation method being up to ten times more accurate than pruning and one-hundred times more accurate than perturbation. The second DSGE model we study is a matching model from Den Haan and De Wind (2012). Here transformed perturbation outperforms pruning up to a factor ten on path errors and a factor thirty for sample moments. Moreover perturbation has path errors that are up to twenty-five times larger and sample moment errors that are up to one-hundred times larger compared to transformed perturbation.

The paper is structured as follows. We start by stating the definition of the transformed perturbation method in Section 2. Section 3 analyses the statistical properties of the transformed perturbation system and provides lenient and accessible conditions that ensure paths are nonexplosive and laws of large numbers can be applied. Section 4 provides a theoretical foundation and motivation for the transformed perturbation approximation method. Finally Section 5 discusses the accuracy of the new method. We provide theoretical results that show that transformed perturbation, accuracy wise, matches regular perturbation locally and strongly outperforms it globally. Moreover, we demonstrate for two example models that transformed perturbation outperforms pruning and regular perturbation on numerous common criteria.

2 Transformed Perturbation

2.1 The state space

Let $\bar{\mathbf{y}}_t$ be an n_y-dimensional vector of control variables, let $\bar{\mathbf{x}}_t$ be an n_x-dimensional vector of endogenous state variables and let $\mathbf{z}_t$ be an n_z-dimensional vector of exogenous state variables. We study the general class of DSGE models characterized by a set of first-order dynamic optimality conditions that can be written as

$$0 = \mathbb{E}_t(f(\bar{\mathbf{y}}_{t+1}, \bar{\mathbf{y}}_t, \bar{\mathbf{x}}_{t+1}, \bar{\mathbf{x}}_t, \mathbf{z}_{t+1}, \mathbf{z}_t)), \tag{1}$$

$$\mathbf{z}_{t+1} = \mathbf{\Lambda}\mathbf{z}_t + \sigma\boldsymbol{\eta}\boldsymbol{\varepsilon}_{t+1}. \tag{2}$$

Here $\mathbb{E}_t$ denotes the expectation operator conditional on the information at time t, and $f : \mathbb{R}^{2(n_x+n_y+n_z)} \to \mathbb{R}^{n_y+n_x}$ is a real function. The matrix $\mathbf{\Lambda}$ is assumed to be invertible with spectral radius smaller than one. Finally σ is the auxiliary perturbation parameter and $\boldsymbol{\varepsilon}_{t+1}$ is a n_z-dimensional vector of exogenous innovations with mean zero and finite second moment that takes values in $\mathcal{E} \subseteq \mathbb{R}^{n_z}$. Throughout the paper, we will assume that $(\boldsymbol{\varepsilon}_t)_{t\in\mathbb{N}}$ is an independent and identically distributed (iid) stochastic process.

We define the deterministic steady states $\mathbf{y}_{ss}$ and $\mathbf{x}_{ss}$ of $\bar{\mathbf{y}}_t$ and $\bar{\mathbf{x}}_t$ respectively such that

$$f(\mathbf{y}_{ss}, \mathbf{y}_{ss}, \mathbf{x}_{ss}, \mathbf{x}_{ss}, \mathbf{0}_{n_z}, \mathbf{0}_{n_z}) = 0.$$

Furthermore, let $\mathbf{y}_t = \bar{\mathbf{y}}_t - \mathbf{y}_{ss}$ and $\mathbf{x}_t = \bar{\mathbf{x}}_t - \mathbf{x}_{ss}$ denote the random variables in deviations from the steady-state, where $\mathbf{y}_t$ takes values in $\mathcal{Y} \subseteq \mathbb{R}^{n_y}$ and $\mathbf{x}_t$ takes values in $\mathcal{X} \subseteq \mathbb{R}^{n_x}$. We write $\mathcal{Z} \subseteq \mathbb{R}^{n_z}$ for the domain of $\mathbf{z}_t$. Following Den Haan and De Wind (2012), the solution to the model given in equation (1) is of the form

$$\mathbf{y}_{t+1} = g(\mathbf{x}_t, \mathbf{z}_{t+1}, \sigma), \tag{3}$$

$$\mathbf{x}_{t+1} = h(\mathbf{x}_t, \mathbf{z}_{t+1}, \sigma). \tag{4}$$

We refer to (3) and (4) as the observation and state equations respectively. It follows from our setup that $g(\mathbf{0}_{n_x}, \mathbf{0}_{n_z}, 0) = \mathbf{0}_{n_y}$ and $h(\mathbf{0}_{n_x}, \mathbf{0}_{n_z}, 0) = \mathbf{0}_{n_x}$.

Both functions g and h, known as *policy functions*, are unknown functions that must be approximated. If the function g in the observation equation is measurable, then the stability of the solution of a DSGE model depends entirely on the state equation. For this reason we will focus on approximating the function h in (4).

2.2 Function Approximation Methods

A wide range of techniques have been proposed in the literature to approximate the unknown policy function h. In most cases, the approximate policy function is obtained as an element of a vector space spanned by a set of basis functions $\{\phi_1, ..., \phi_m\}$:

$$h(\mathbf{x}, \mathbf{z}, \sigma) \approx \sum_{i=1}^{m} A_i \phi_i(\mathbf{x}, \mathbf{z}, \sigma),$$

where $A_1, ..., A_m$ are matrices of coefficients that weight the basis functions $\phi_1, ..., \phi_m$.

There exist a multitude of popular sets of basis functions and weight matrix calculation methods that have been proposed in the function approximation literature. Well known classes of basis functions include *power monomials*, which are used with great success in Taylor expansions, *sigmoid trigonometric functions*, that are prominently featured in Fourier approximations, *Chebyshev polynomials*, that play an important role on orthogonal polynomial function approximation, *Legendre polynomials*, which are often used for approximating density functions, and *logistic functions*, comprehensively explored in artificial neural network approximations. Popular methods for calculating the weight matrices include *Taylor's method* which obtains the matrices as weighted derivatives at a given expansion point and minimizes the so called Taylor semi-norm (Apostol, 1967), *function colocation methods*, which minimize a discrete distance between the true and approximate policy function at a finite number of points and *spectral approximation methods*, that minimize a continuous distance between the two functions. See e.g. Powell (1981) for an overview of approximation literature and Judd (1998) for an application of these methods to approximating policy functions of dynamic stochastic models.

2.3 Perturbation

Perturbation is a method that approximates the unknown policy function h by using power monomials as basis functions in combination with Taylor's method to find the weighting matrices. This method is of particular interest in approximating policy functions of DSGE models as it provides a fast and analytically tractable way of obtaining the weighting matrices. The expansion point used in Taylor's method is the deterministic steady state $(\mathbf{0}_{n_x}, \mathbf{0}_{n_z}, 0)$. Choose $\mathbf{x} \in \mathcal{X}$ and $\mathbf{z} \in \mathcal{Z}$ and define $\mathbf{v} = (\mathbf{x}, \mathbf{z})$ and $\bigotimes_i \mathbf{v} = \underbrace{\mathbf{v} \otimes \cdots \otimes \mathbf{v}}_{i \text{ times}}$, where the empty Kronecker product is set to one. Then the m'th order perturbation approximation of h evaluated at $(\mathbf{x}, \mathbf{z}, \sigma)$ can be expressed as

$$h_p(\mathbf{x}, \mathbf{z}, \sigma) := H_0 + H_{\mathbf{x}}\mathbf{x} + H_{\mathbf{z}}\mathbf{z} + \sum_{i=2}^{m} H_i \bigotimes_i \mathbf{v}, \tag{5}$$

where we grouped all terms of $\mathbf{v}$ of the same power, regarding σ as a constant. That is,

$$H_0 = \sum_{j=0}^{m} \frac{1}{j!} \frac{\partial^j}{\partial \sigma^j} h(\mathbf{0}_{n_x}, \mathbf{0}_{n_z}, 0)\sigma^j \qquad H_{\mathbf{x}} = \sum_{j=0}^{m-1} \frac{1}{j!} \frac{\partial^{j+1}}{\partial \sigma^j \partial \mathbf{x}} h(\mathbf{0}_{n_x}, \mathbf{0}_{n_z}, 0)\sigma^j$$

$$H_{\mathbf{z}} = \sum_{j=0}^{m-1} \frac{1}{j!} \frac{\partial^{j+1}}{\partial \sigma^j \partial \mathbf{z}} h(\mathbf{0}_{n_x}, \mathbf{0}_{n_z}, 0)\sigma^j \qquad H_i = \sum_{j=0}^{m-i} \frac{1}{i!j!} \frac{\partial^{j+i}}{\partial \sigma^j \partial \mathbf{v}^i} h(\mathbf{0}_{n_x}, \mathbf{0}_{n_z}, 0)\sigma^j$$

Thus, H_0 is an $n_x \times 1$ vector that is the sum of all the derivatives of h with respect to powers of σ. The matrix $H_{\mathbf{x}}$ is an $n_x \times n_x$ matrix that is the sum of all the derivatives of h with respect to $\mathbf{x}$ and powers of σ. The matrix $H_{\mathbf{z}}$ is an $n_x \times n_z$ matrix that is the sum of all the derivatives of h with respect to $\mathbf{z}$ and powers of σ. Finally, the matrices H_i are of dimension $n_x \times (n_x + n_z)^i$ and given by the sum of all the derivatives of h with respect to $\mathbf{v}^i$ and powers of σ.

2.4 The transformed perturbation method

A disadvantage of the power monomial set of basis functions, and therefore of perturbation, is that the derivative of the approximation function tends to infinity away from the steady state if $m > 1$. This creates highly explosive regions in the state space which in practice means that sample paths eventually diverge to infinity with probability one. The transformed perturbation method solves this problem by using another set of basis functions called the transformed power monomials. This set of basis functions satisfies all the advantageous properties that classical power monomials do. Blasques (2014) shows that transformed polynomials with unrestricted weighting matrices can be used to approximate continuous functions with arbitrary accuracy, in the same way as classical polynomials, by application of the *Stone-Weierstrass Theorem* (Stone, 1937, 1948). Additionally, Blasques (2014) characterizes the convergence rates of transformed polynomials on Sobolev spaces of smooth n-times continuously differentiable functions with nth derivative bounded in L_p norm, through the application of Plesniak's extension of *Jackson's Theorem* (Plesniak, 1990).

The set of transformed power monomials is obtained by multiplying the monomials of order greater than one with an exponentially fast decaying function $\Phi_\tau : \mathcal{X} \to \mathbb{R}$ that is a multivariate adaptation of the transformed function of Blasques (2014) and is defined as

$$\Phi_\tau(\mathbf{x}) = e^{-\tau\|\mathbf{x}\|_e^2}, \tag{6}$$

where $\|\mathbf{x}\|_e$ denotes the Euclidean norm of $\mathbf{x}$. Figure 1 plots the second and third order one dimensional transformed monomials for varying values of τ. Note that the case τ is zero

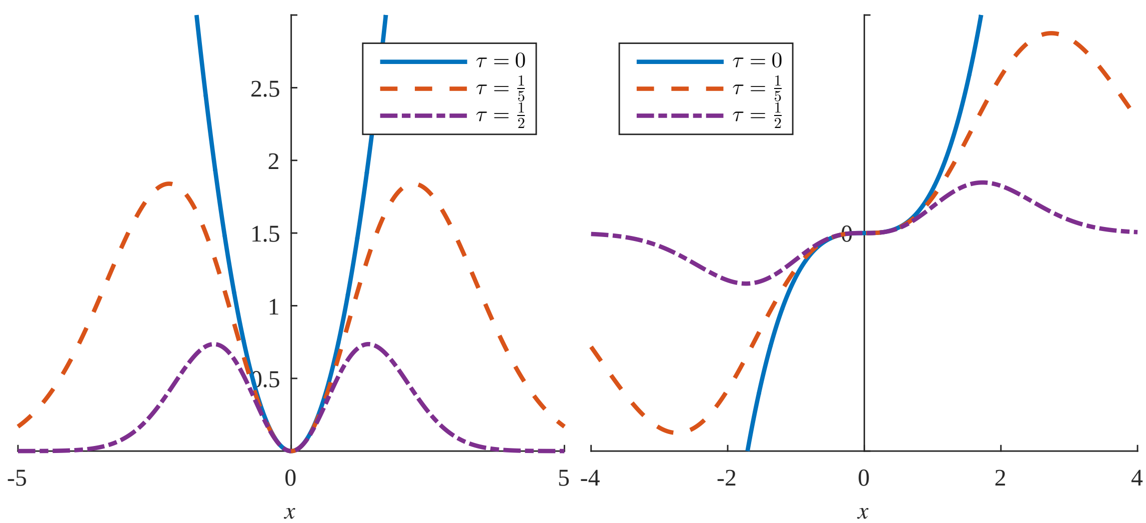


Figure 1: Plots of the second, respective third, order one dimensional transformed monomial in the left, respective right, panel for values of $\tau \in \{0, 0.2, 0.5\}$.

sets the transformed monomials equal to the regular monomial basis functions. The figure shows that the transformed monomials are almost identical to regular monomials close to the steady state at zero. However, the derivatives of transformed monomials vanish away from the steady state, which implies that no explosive regions are created in the state space. In Section 5 we will further show that transformed perturbation has the same local approximation properties as classical perturbation. In particular, local approximation rates are the same as for classical perturbation, and transformed perturbation approximations converge uniformly on compact analytic domains, just like perturbation methods do. A large number of additional advantages of transformed perturbation over classical perturbation and pruning methods are documented in Section 3 and Section 5.

State variables can be of different orders in size, so we replace the vector $\mathbf{x}$ in (6) by the relative differences from the steady state $\tilde{\mathbf{x}} = \mathbf{x}/\mathbf{x}_{ss}$, where dividing is done entry wise, to ensure all variables have equal effect. This definition works poorly if an entry of $\mathbf{x}_{ss}$ is close to zero. For such an entry we take the simple transformation $x \mapsto e^x \approx 1 + x$, which is almost linear close to zero, and define $\tilde{x} = (e^{x+x_{ss}} - e^{x_{ss}})/e^{x_{ss}}$. The m'th order transformed perturbation approximation of h evaluated at $(\mathbf{x}, \mathbf{z}, \sigma)$ is then defined as

$$h_{tp}(\mathbf{x}, \mathbf{z}, \sigma) = H_0 + H_{\mathbf{x}}\mathbf{x} + H_{\mathbf{z}}\mathbf{z} + \left(\sum_{i=2}^{m} H_i \bigotimes_i \mathbf{v}\right) \Phi_\tau(\tilde{\mathbf{x}}), \tag{7}$$

where all the H matrices are obtained using Taylor's method and thus they are identical to those in the regular perturbation function (5).

The constant τ determines the speed at which the higher order terms in (7) are going to zero when moving away from the origin. Its value influences the shape of the resulting policy function, and thus requires careful consideration. We offer two methods to set τ. The first method is to find the optimal τ, denoted τ^*, by minimizing some criterion function. In this paper we chose to minimise the maximum Euler errors on a relevant set around the steady state. The advantage of this method is that we get the best possible value for τ, according to the criterion function. The disadvantage is that minimizing the criterion function potentially is time-consuming. In an estimation setting we fix the optimal τ^* at the start and then estimate the remaining parameters while τ^* remains fixed. This means that the possibly time consuming task of finding τ^* has to be executed only once, making the method almost as fast as perturbation, still viable for estimation and very accurate if the optimal τ^* does not vary too much with the parameters. The second method is designed to avoid the optimization completely and is characterised by a plug-in τ, denoted $\hat{\tau}$, which is less precise, but found immediately. The plug-in value is given by

$$\hat{\tau} = \frac{1}{c} \log\left(\frac{1}{1 - \rho(H_{\mathbf{x}})}\right), \tag{8}$$

where $\rho(H_{\mathbf{x}})$ is the spectral value of the autoregressive part of the regular perturbation solution and c is an approximation of the average range that the state variables take place in. This range could be set according to prior knowledge on the variables, or approximated by another solution method. In our case we used linear perturbation to simulate a series and find the approximate range of our variables. In an estimation setting we update $\hat{\tau}$ as the parameters are updated, since its calculation is very fast. See Section 4 for a detailed discussion on the choice for our plug-in value.

3 Probabilistic analysis of the solutions

Throughout this paper we work with norms $\|\cdot\|$ on Euclidean space and their induced matrix norms, which we will denote with the same notation $\|\cdot\|$ as there should be no confusion in their use. Note that all matrix norms are equivalent, so that our statements will work for any chosen norm.

Let $\mathbf{x}_0 \in \mathcal{X}$ and $\mathbf{z}_0 \in \mathcal{Z}$ be fixed and define the exogenous sample paths $(\mathbf{z}_t)_{t\geq 0}$ and the transformed perturbation sample paths $(\mathbf{x}_t)_{t\geq 0}$ recursively by

$$\begin{aligned}\mathbf{z}_{t+1} &= \boldsymbol{\Lambda}\mathbf{z}_t + \sigma\boldsymbol{\eta}\boldsymbol{\varepsilon}_{t+1},\\ \mathbf{x}_{t+1} &= h_{tp}(\mathbf{x}_t, \mathbf{z}_{t+1}, \sigma).\end{aligned}$$

In this section we analyse the dynamics of the transformed perturbation system and provide two results on the stability of sample paths. To do so we split the perturbation updating equation (7) into the sum of its linear part $H_0 + H_{\mathbf{x}}\mathbf{x} + H_{\mathbf{z}}\mathbf{z}$ and its nonlinear part

$$D(\mathbf{x}, \mathbf{z}) := \left(\sum_{i=2}^{m} H_i \bigotimes_i \mathbf{v} \right) \Phi_\tau(\tilde{\mathbf{x}}). \tag{9}$$

Our results are based on the observation that the transformed perturbation policy function (7) is asymptotically equal to its linear part as $\|\mathbf{x}\| \to \infty$. This follows because an exponential function decays at greater speed than a polynomial, see Figure 1, and thus for any $0 \le i \le m$ we have

$$\lim_{\|\mathbf{x}\| \to \infty} \left(\bigotimes_i \mathbf{x} \right) \Phi_\tau(\tilde{\mathbf{x}}) = \mathbf{0}_{n_x^i}.$$

We therefore study the transformed perturbation method as its asymptotic linear process plus a deviation (9). Linear autoregressive processes and their stability have been extensively studied. They are much easier to analyse compared to their nonlinear counterparts, because we get an analytical closed form when we expand the expressions for $\mathbf{x}_t$ and $\mathbf{z}_t$ back in time. It can be shown that if backwards expanding converges, then the limit is a stationary ergodic solution to the system. See Theorem 3.1 in Bougerol (1993) for a general result on the stability of contracting systems that uses this approach. In our first result we closely mimic this technique by bounding the deviation from the linear process. We require the following assumptions.

Assumption A.

A1. The spectral radius $\rho(\mathbf{\Lambda}) < 1$.

A2. The spectral radius $\rho(H_{\mathbf{x}}) < 1$.

A3. There exists an $r > 0$ such that $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^{rm} < \infty$.

Our first result shows that solution paths generated by the transformed perturbation solution are non-explosive almost surely if Assumption A holds. The conditions in Assumption A are very lenient. Assumption A2 is close to being both sufficient and necessary. The spectral radius is a measure for the maximal scale at which $H_{\mathbf{x}}$ can stretch a vector. Therefore, if $\rho(H_{\mathbf{x}}) > 1$, then an eigenvector belonging to the eigenvalue that is greater than one in absolute value is expanded by $H_{\mathbf{x}}$. If the space spanned by this vector is reachable from the exogenous variable space $\mathcal{Z}$, then expanding backwards will explode and thus diverge. Assumption A3 is satisfied for any r if, for example, the $\boldsymbol{\varepsilon}_t$ have finite support, or are normally distributed, or have sub-exponential tails. Additionally, for fat tailed distributions, the moments of $\mathbf{x}_t$ and $\mathbf{z}_t$ are a fraction of those of the innovations.

Theorem 3.1 (Non explosive paths). *Suppose that Assumption A holds. Then the dynamic system defined in* (2) *and* (4)*, featuring the transformed perturbation policy function given in* (7)*, produces sample paths that are non explosive almost surely, i.e. the paths* $(\mathbf{z}_t)_{t\in\mathbb{N}}$ *and* $(\mathbf{x}_t)_{t\in\mathbb{N}}$ *satisfy*

$$\liminf_{t\to\infty} \|\mathbf{z}_t\| < \infty \qquad \textit{and} \qquad \liminf_{t\to\infty} \|\mathbf{x}_t\| < \infty \qquad \textit{a.s.}$$

Theorem 3.1 shows that the transformed perturbation method does not produce explosive paths, unlike regular perturbation sample paths. However, we can show much more. Our stability results are based on Markov chain theory as developed in Meyn and Tweedie (1993). We are in a Markov chain setting, because we have assumed that $(\boldsymbol{\varepsilon}_t)_{t\in\mathbb{N}}$ is an iid sequence. We provide two sets of assumptions, the first of which is more general and harder to verify, while the second set imposes additional constraints that are straightforward to verify.

A point $\mathbf{x}^* \in \mathcal{X}$ is called *reachable* if for every open set $\mathbf{x}^* \in O \subseteq \mathcal{X}$ and starting value $\mathbf{x}_0 \in \mathcal{X}$ there exists a $t \in \mathbb{N}$ such that $\mathbb{P}(\mathbf{x}_t \in O) > 0$. A subset of $\mathcal{X}$ is called reachable if all the points in it are reachable. We will need the following additional assumptions.

Assumption B.

B1. $\mathcal{X}$ has an open reachable subset.

B2. The innovation $\boldsymbol{\varepsilon}_t$ is absolutely continuous with respect to the Lebesque measure on $\mathcal{E}$ with strictly positive density on a connected subset of $\mathcal{E}$.

Our second result establishes the stationarity and ergodicity of the transformed perturbation solution. Additionally, it shows that the solution paths have fading memory in the sense of geometric ergodicity and absolutely regularity (or β-mixing) of the process. Finally, it also shows that the solution paths have finite r-th moment. Stationarity, fading memory, and bounded moments are all important ingredients in the statistical analysis of DSGE models, from estimation to probabilistic analysis.

Theorem 3.2. (Stationarity, fading-memory and bounded moments) *Suppose that Assumptions A and B hold. Then there exists a unique stationary ergodic solution* $(\mathbf{x}_t^*, \mathbf{z}_t^*)_{t\geq 0}$ *to the dynamic system defined in* (2) *and* (4)*, featuring the transformed perturbation policy function given in* (7)*. Additionally,*

(i) the solution has fading memory, i.e. it is geometrically ergodic and absolutely regular (or β-mixing);

(ii) the solution has finite moments $\mu_r := \mathbb{E}\|\mathbf{x}_t^*\|^r$ *and* $\nu_{rm} = \mathbb{E}\|\mathbf{z}_t^*\|^{rm}$*;*

(iii) laws of large numbers apply to the sample paths, that is, almost surely

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{x}_t\|^r = \mu_r \qquad \textit{and} \qquad \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \|\mathbf{z}_t\|^{rm} = \nu_{rm}.$$

Assumption B imposes additional conditions on our state space system. Assumption B2 is quite weak and is satisfied for all distributions that are used in practice. The stronger, and also harder to check, condition is Assumption B1. We present Assumption B, because simplifying Assumption B1 will require us to assume that the innovations have full support. This is not always the case, as we might, for example, have strictly positive innovations. If we can make the assumption of full support, then we get an easier set of conditions.

Assumption C.

C1. There exists an integer $t \geq 1$ such that the matrix $\begin{bmatrix} H_{\mathbf{x}}^{t-1}H_{\mathbf{z}} & \cdots & H_{\mathbf{x}}H_{\mathbf{z}} & H_{\mathbf{z}} \end{bmatrix}$ has rank n_x.

C2. The matrix $H_{\mathbf{x}}$ is invertible.

C3. The innovation $\boldsymbol{\varepsilon}_t$ is absolutely continuous with respect to the Lebesque measure on $\mathbb{R}^{n_z}$ with strict positive density on the whole space $\mathbb{R}^{n_z}$.

Proposition 3.3. *Assumption C implies Assumption B.*

Assumption C2 ensures that the transformed perturbation policy function does not move to lower dimensional subspaces of $\mathcal{X}$. Condition C1 implies that the effect of the innovations is not contained in a lower dimensional subspace. This means that, together with Assumption C3, they make sure that the transformed perturbation policy function can reach any point in $\mathcal{X}$ and thus Assumption B1 is satisfied.

4 The plug-in $\hat{\tau}$

In this section we motivate our choice for $\hat{\tau}$, the plug in value of τ, as defined in (8). As mentioned in Section 2, its value influences the shape of the transformed perturbation policy function and thus has an effect on sample path behaviour in the resulting transformed perturbation dynamic system. We want to ensure two important properties for this dynamic system. Firstly, we want sample paths to be stable and non locally explosive. In Section 4.1 we argue that this requires relatively large values of τ. Secondly, nonlinear dynamics must be preserved, which needs τ to take on somewhat small values, see Section 4.2. Together these two conditions specify a rather narrow collection of available functions, resulting in (8), as derived in Section 4.3.

4.1 Ensuring stability

The transformed perturbation method guarantees stable and nonexplosive paths regardless of the choice of τ, as proved in Section 3. However, picking τ very small can create locally explosive dynamics. Locally explosive dynamics originate when the jacobian of the policy function with respect to $\mathbf{x}$ has expected spectral radius greater than one on a large enough subset of $\mathcal{X}$. A spectral radius greater than one implies that the policy function expands on some subspace, which can create multiple fixed points, as happens with the regular perturbation policy function. Sample paths then typically move around one fixed point, until a large innovation pushes it to another fixed point after which the path moves around the new one. These jumps can locally look very similar to explosive sample paths, even though the dynamic system is stable. We illustrate this effect with the following example updating equation

$$x_{t+1} = 0.3x_t + z_{t+1} + 2x_t^3 e^{-0.5x_t^2}, \tag{10}$$

where the $(z_t)_{t\in\mathbb{N}}$ are updated as in (2). Note that this is a univariate example of (7) with $\tau = 0.5$. Figure 2a plots the expected value $\mathbb{E}(x_{t+1} \mid x_t)$ as a function of x_t. This function has large intervals on which its absolute derivative exceeds one, which has resulted in a total of five fixed points. The smallest one at -2.35, the middle one at zero and the largest one at 2.35 are attractors while the other two are repellers. A sample path produced while using (10) will jump between the neighbourhoods around the three attractors. Figure 2b plots an example sample paths that first spends some time around -2.35, then jumps to a neighbourhood of the origin and then quickly moves on to the area around the largest

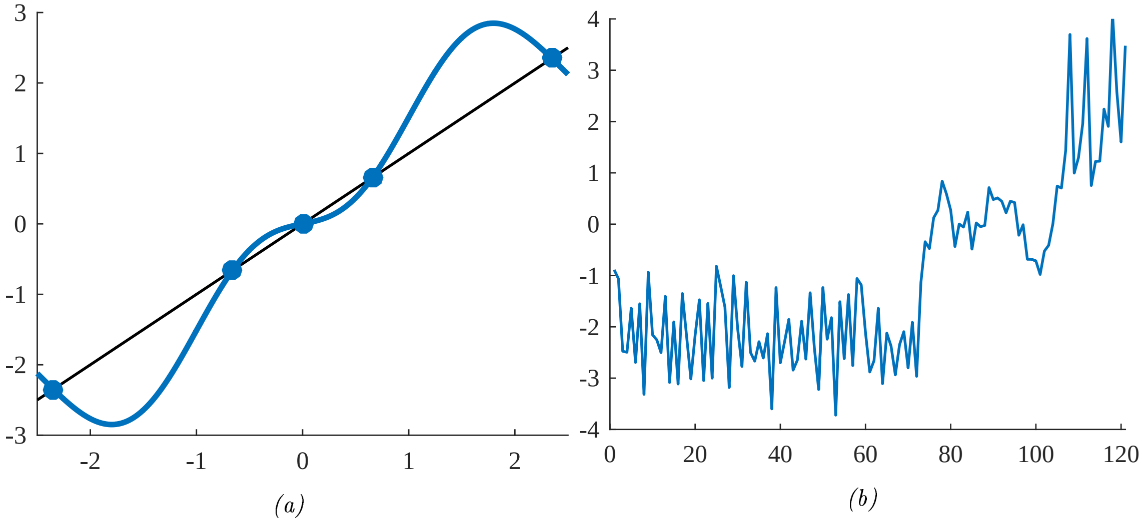


Figure 2: The expected policy function (left panel) and an example sample path (right panel) for the updating equations defined in (10) *and* (2)*.*

attractor. Notice the similarity with an explosive sample path, even though this path will almost surely eventually come down to the lowest attractor again.

We wish to keep the spectral value of the Jacobian of the transformed perturbation policy function typically below one (in expectation) to avoid locally explosive behaviour. This Jacobian is of the form

$$J = H_{\mathbf{x}} + P(\mathbf{x}, \mathbf{z})\Phi_\tau(\tilde{\mathbf{x}}),$$

where P is a m'th order multivariate polynomial function. We can only control the nonlinear part of the derivative, i.e. the second part of the summation, with our choice for τ. Any norm of P goes to infinity as $\|\mathbf{x}\|$ goes to infinity. Hence, if we choose τ too small, then $P(\mathbf{x}, \mathbf{z})\Phi_\tau(\tilde{\mathbf{x}})$ creates large areas on the state space with expected spectral radius greater than one. If we were only concerned with ensuring stability, then ideally we would choose $\tau = \infty$, so that Assumption A2 ensures that $\rho(J) < 1$ on the entire state space. Doing so, however, cancels all nonlinear effects making the transformed perturbation method equal to linear perturbation, which as discussed in the introduction has many flaws. Therefore we conclude that we would like to make τ as large as possible, while preserving as much nonlinear dynamics as possible close to the steady state. If we choose τ unequal to infinity, then its size generally must depend on $\rho(H_{\mathbf{x}})$. The closer $\rho(H_{\mathbf{x}})$ is to one, the less room remains available for $P(\mathbf{x}, \mathbf{z})\Phi_\tau(\tilde{\mathbf{x}})$. Accordingly we have to impose that τ goes to infinity as $\rho(H_{\mathbf{x}})$ gets closer to one. Therefore we must find a function $f : (0, 1) \to [0, \infty)$ such that $\tau = f(\rho(H_{\mathbf{x}}))$ and

$$\lim_{\rho(H_{\mathbf{x}}) \to 1} f(\rho(H_{\mathbf{x}})) = \infty. \tag{11}$$

4.2 Preserving nonlinear dynamics

We have concluded that we want to choose large τ to avoid locally explosive behaviour, but not so large as to destroy relevant nonlinear dynamics. In this section we formalise what we mean with preserving nonlinear dynamics. To do so we expand $\mathbf{x}_t$ back in time, mimic the proof of Theorem 3.1 and use Proposition A.1 to find the upper bound

$$\|\mathbf{x}_t\| \leq \tilde{c} + \sum_{k=0}^{\infty} \|H_{\mathbf{x}}\|^k \|H_{\mathbf{z}}\| \|\mathbf{z}^*_{t-k}\| + c \sum_{j=0}^{m} \sum_{k=0}^{\infty} \|H_{\mathbf{x}}\|^k \tau^{-j/2} \left(\sum_{i=0}^{m-j} \|\mathbf{z}^*_{t-k}\|^i \right).$$

for some constants $c, \tilde{c} > 0$. The first, respective second, summation here is the approximate total effect over time of the linear, respective nonlinear, terms in (7). The first summation

$$\sum_{k=0}^{\infty} \|H_{\mathbf{x}}\|^k \|H_{\mathbf{z}}\| \|\mathbf{z}^*_{t-k}\|,$$

is the familiar term that arises in autoregressive processes.[2] The autoregressive part $H_{\mathbf{x}}\mathbf{x}$ of the policy function (7) introduces memory into the system, so that past innovations $\|\mathbf{z}_{t-k}\|$ influence the value of $\|\mathbf{x}_t\|$. The strength of the memory depends on the size of $\rho(H_{\mathbf{x}})$. If it is close to zero, then memory fades away fast and past innovations are of little weight to $\mathbf{x}_t$. As $\rho(H_{\mathbf{x}})$ increases, past innovations matter more up to the limit case $\rho(H_{\mathbf{x}}) = 1$, where memory does not fade anymore, at which point every past innovation is equally important and the sum diverges for all matrix norms.

We would like the impact of past innovations through the nonlinear terms of the transformed perturbation policy function to be of the same magnitude as those of the linear effect, so that both the linear and nonlinear dynamics are present in the solution paths. Specifically, we want the rate at which τ goes to infinity to be restricted such that the series

$$\sum_{k=0}^{\infty} \|H_{\mathbf{x}}\|^k \tau^{-j/2} \left(\sum_{i=0}^{m-j} \|\mathbf{z}^*_{t-k}\|^i \right)$$

diverge for all $0 \leq j \leq m$ as $\rho(H_{\mathbf{x}}) \to 1$. If this were not the case, then they would converge and thus we would restrict some nonlinear effects so much that the linear effect is infinitely stronger as $\rho(H_{\mathbf{x}})$ increases. To ease notation we define $\delta_t = \sum_{i=0}^{m-j} \|\mathbf{z}^*_t\|^i$. The argument above then amounts to the following desired result: for all $j \in \mathbb{N}$ we have

$$\lim_{\rho(H_{\mathbf{x}}) \to 1} \sum_{k=0}^{\infty} \rho(H_{\mathbf{x}})^k \tau^{-j/2} \delta_{t-k} = \lim_{\rho(H_{\mathbf{x}}) \to 1} f(\rho(H_{\mathbf{x}}))^{-j/2} \sum_{k=0}^{\infty} \rho(H_{\mathbf{x}})^k \delta_{t-k} = \infty. \tag{12}$$

It is not immediately clear what divergence rates for $f(\rho(H_{\mathbf{x}}))$ satisfy (12). Therefore we include the following result to simplify the expression.

Lemma 4.1. *Suppose that $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^m < \infty$. Then the limit*

$$\lim_{\rho(H_{\mathbf{x}}) \to 1} (1 - \rho(H_{\mathbf{x}})) \sum_{k=0}^{\infty} \rho(H_{\mathbf{x}})^k \delta_{t-k}$$

converges to a finite and nonzero value.

It now follows from Lemma 4.1 that (12) is equivalent to

$$\lim_{\rho(H_{\mathbf{x}}) \to 1} f(\rho(H_{\mathbf{x}}))^{j/2} (1 - \rho(H_{\mathbf{x}})) = 0. \tag{13}$$

[2] Note that it converges by Assumption A, Proposition 2.5.1 of Straumann (2005) and Proposition 4.3 of Krengel (1985).

4.3 Choice for $\hat{\tau}$

We need a function $f : (0,1) \to (0,\infty)$ that satisfies both (11) and (13). To simplify these equations further we define $\tilde{f} : (0,\infty) \to [0,\infty)$ as $\tilde{f}(\frac{1}{1-\rho(H_{\mathbf{x}}))}) = f(\rho(H_{\mathbf{x}}))$ and substitute $y = \frac{1}{1-\rho(H_{\mathbf{x}}))}$. Equations (11) and (13) then can be rewritten as

$$\lim_{y\to\infty} \tilde{f}(y) = \infty \qquad \text{and} \qquad \lim_{y\to\infty} \frac{\tilde{f}(y)^{j/2}}{y} = 0.$$

These two equations together specify a fairly small collection of functions. To find the function that diverges fastest we consider families of familiar functions in decreasing order of rate of divergence. Note that any exponential, polynomial or radical function diverges to infinity too fast to satisfy the rightmost limit for all $j \in \mathbb{N}$. The next natural candidate in line for the rate of divergence is the logarithmic function, which leads to the specification

$$f(\rho(H_{\mathbf{x}})) = \log\left(\frac{1}{1-\rho(H_{\mathbf{x}}))}\right).$$

This is the function we used for our choice in (8).

The constant τ should also depend on the size of the range on which the state variables take place. Suppose that we increase the scale of our dynamic system while keeping the exact same dynamics. Then τ should become smaller as regions farther away from the steady-state are visited more often. Therefore we include the c parameter to make sure that as we make the scale larger, τ becomes smaller. Many of the other elements involved in the perturbation updating function, such as σ or H_i for $i \geq 2$ seem to be omitted in calculating the plug in τ. However, these elements have an effect on the range of the state variables and thus are implicitly included via c.

5 Accuracy

In this section we evaluate the accuracy of the transformed perturbation solution. In Section 5.1 we prove theoretic results on both global and local accuracy. We show that the optimal transformed perturbation solution is always at least as accurate as regular perturbation and demonstrate that transformed polynomials, like regular polynomials can perfectly approximate the real policy function h as we let the approximation order m go to infinity. Moreover, we prove that transformed perturbation is locally as accurate as standard perturbation and present common situations in which transformed perturbation globally outperforms regular perturbation. Section 5.2 discusses two DSGE models from Den Haan and De Wind (2012) and compares all discussed solution methods according to several criteria such as path errors, euler errors and produced moments. It shows that transformed perturbation outperforms pruning and regular perturbation for both the optimal τ^* and the plug in $\hat{\tau}$.

5.1 Theoretical results

In order to analyse the accuracy of our approximation method we define the pointwise approximation errors attained by the perturbation and transformed perturbation methods respectively, at $(\mathbf{x}, \mathbf{z}, \sigma) \in \mathcal{X} \times \mathcal{Z} \times \mathbb{R}_{\geq 0}$ as

$$E_p(\mathbf{x}, \mathbf{z}, \sigma) := \|h_p(\mathbf{x}, \mathbf{z}, \sigma) - h(\mathbf{x}, \mathbf{z}, \sigma)\|,$$
$$E_{tp}(\mathbf{x}, \mathbf{z}, \sigma) := \|h_{tp}(\mathbf{x}, \mathbf{z}, \sigma) - h(\mathbf{x}, \mathbf{z}, \sigma)\|.$$

We begin by showing that the function approximation by transformed perturbation converges on analytic domains, like the standard perturbation approximation. This result implies that we can arbitrarily accurately approximate the true policy function by increasing the order m.

Proposition 5.1. *Suppose that the true policy function is analytic over a compact set $S \subseteq \mathcal{X} \times \mathcal{Z} \times \mathbb{R}_{\geq 0}$. Then m-order transformed perturbation errors vanish uniformly over S for any sequence $\tau \to 0$ as the perturbation order diverges to infinity. That is,*

$$\lim_{m\to\infty,\tau\to 0} \sup_{(\mathbf{x},\mathbf{z},\sigma)\in S} E_{tp}^{(m)}(\mathbf{x},\mathbf{z},\sigma) := \|h_{tp}^{(m)}(\mathbf{x},\mathbf{z},\sigma) - h(\mathbf{x},\mathbf{z},\sigma)\| = 0.$$

Next, we prove that transformed perturbation is always able to outperform regular perturbation.

Proposition 5.2. *For any policy function h there exists a $\tau \geq 0$ such that $E_{tp}(\mathbf{x},\mathbf{z},\sigma) \leq E_p(\mathbf{x},\mathbf{z},\sigma)$ for all possible values of $\mathbf{x}$, $\mathbf{z}$ and σ.*

Note that this result makes no assumptions on the true policy function and implies that using the optimal τ^* for the transformation guarantees an equal or better approximation compared to regular perturbation. This result is true even when regular perturbation sample paths do not seem to explode. Therefore, it may be argued that transformed perturbation should always be used over regular perturbation.

We proceed by studying the accuracy properties of the transformed perturbation method for arbitrary values of the constant τ. First we show that locally the transformed polynomials inherit the excellent approximation qualities of perturbation methods. This follows because the exponential function $\Phi_\tau(\tilde{\mathbf{x}})$ is asymptotically quadratic as $\|\mathbf{x}\|$ goes to zero. A consequence of the proposition below is that, close to the steady state, errors between the transformed perturbation paths and the true paths are of the same magnitude as the errors between the regular perturbation paths and the true paths for $m = 2, 3$.

Proposition 5.3. *Suppose that $\mathbf{x}_0 = \mathbf{0}_{n_x}$ and $\mathbf{z}_0 = \mathbf{0}_{n_z}$. Let $(\mathbf{x}_t)_{t\geq 0}$ be the path generated by the true policy function* (4) *and let $(\hat{\mathbf{x}}_t)_{t\geq 0}$ be the path generated by the m'th order transformed perturbation policy function, both initialised at these same starting values. Then it holds for all $t \in \mathbb{N}$ that*

$$\|\hat{\mathbf{x}}_t - \mathbf{x}_t\| = \begin{cases} O\left(\sigma^3\right) & \text{if } m = 2 \\ O\left(\sigma^4\right) & \text{if } m > 2 \end{cases} \qquad \text{as} \qquad \sigma \to 0.$$

Transformed perturbation has the same local properties as regular perturbation, but on a global scale it is almost guaranteed to perform much better. Clearly if the true policy function produces explosive sample paths, then our method, which does not, cannot be assured to work well. The next result exhibits a very general set up in which the true policy function is ensured to produce nonexplosive sample paths making transformed perturbation infinitely more accurate in the tails than regular perturbation.

Proposition 5.4. *Suppose that Assumptions A1, A3 and C3 hold and that the true policy function h satisfies*

$$\limsup_{\|\mathbf{x}\|\to\infty} \frac{\mathbb{E}(\|h(\mathbf{x},\mathbf{z}_1,\sigma)\| \mid \mathbf{z}_0 = \mathbf{z})}{\|\mathbf{x}\|} < 1 \tag{14}$$

for all possible values of $\mathbf{z}$ *and* σ*. Then the true policy function almost surely produces nonexplosive sample paths and if* $h_p(\mathbf{x}, \mathbf{z}, \sigma)$ *contains a nonzero higher order monomial in* $\mathbf{x}$*, then*

$$\lim_{\|\mathbf{x}\| \to \infty} \frac{E_{tp}(\mathbf{x}, \mathbf{z}, \sigma)}{E_p(\mathbf{x}, \mathbf{z}, \sigma)} = 0 \tag{15}$$

for all possible values of $\mathbf{z}$ *and* σ *outside of a set of Lebesque measure zero.*

Moreover, condition (14) *is implied by each of the following common conditions that are found in the literature on stable stochastic dynamic systems. The true policy function* h

(i) is eventually bounded by the 45 degree line for all possible values of $\mathbf{z}$ *and* σ*. That is,*

$$\limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h(\mathbf{x}, \mathbf{z}, \sigma)\|}{\|\mathbf{x}\|} < 1.$$

(ii) is uniformly contracting for all possible values of $\mathbf{z}$ *and* σ*. That is,*

$$\sup_{\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}} \frac{\|h(\mathbf{x}_1, \mathbf{z}, \sigma) - h(\mathbf{x}_2, \mathbf{z}, \sigma)\|}{\|\mathbf{x}_1 - \mathbf{x}_2\|} < 1.$$

(iii) is slowly varying at infinity for all possible values of $\mathbf{z}$ *and* σ*. That is, for all* $a > 0$ *we have*

$$\lim_{\|\mathbf{x}\| \to \infty} \frac{h(a\mathbf{x}, \mathbf{z}, \sigma)}{h(\mathbf{x}, \mathbf{z}, \sigma)} = 1.$$

5.2 Applications

In this section, we revisit two DSGE models used in Den Haan and De Wind (2012) to compare transformed perturbation to pruning and other solution methods. Below, we will show that the transformed perturbation approximation significantly outperforms both the regular perturbation approximation and the pruning method. For the purpose of comparing the performance of different solution methods, the true policy function will be approximated to an arbitrary level of accuracy on a relevant set using techniques such as projection methods or value function iteration, see Aruoba et al. (2006). We can then compare the solution methods by analysing sample paths between the “true" solutions and the approximated ones. The length of our time paths are $T = 10^4$, with a burn in period of 500 observations.

We compare sample paths according to three different criteria. The first one measures the distance between a period t variable generated by an approximation versus the one generated by the true policy function as in Den Haan and De Wind (2012). Let x_t be a generalisation of a univariate variable according to the true solution, let $\dot{x}_t$ be generated according to some approximation and let M be the mean of the path $(x_t)_{t=1}^T$. Then we define the error at time t as

$$\min\left\{\left|\frac{\dot{x}_t - x_t}{x_t}\right|, \left|\frac{\dot{x}_t - x_t}{M}\right|\right\},$$

that is, we take the minimum of the absolute percentage error and the absolute error relative to the mean of the true solution path. The minimum between these two is chosen because the percentage error inflates the error when x_t is close to zero, while the error scaled by the mean overestimates inaccuracy when variables take on values far away from their mean.

The second criteria that we use are Euler errors. The equilibrium condition (1) is typically unequal to zero when we use an approximation method instead of the true solution. Its size is an indication for accuracy, because the size of the difference in supremum norm on a compact set between an approximate policy function and the true solution is of the same magnitude as the Euler error, see Theorem 3.3 of Santos (2000). We report the non normalized sample Euler error. We don't normalize our Euler errors, because we are only interested in relative accuracy.

Finally we compare sample moments generated by the approximated paths versus the true ones. DSGE models are often estimated using moment based approaches such as the (simulated) method of moments or indirect inference. Therefore the accuracy of the moments will have an impact on the estimated parameters. Let x_t and y_t be univariate variables, then we compare the sample moments

$$\mu^k(x_t) = \frac{1}{T}\sum_{t=1}^{T} x_t^k$$

and cross moments $\mu(x_t^i y_t^j)$.

5.2.1 The Deaton model

The first model we consider is a partial equilibrium model in which agents face idiosyncratic income risk. The original model was proposed in Deaton Angus (1991), however, we use the modified penalty function that was introduced in Den Haan and De Wind (2012) to compare pruned and non-pruned perturbation solution methods. The model is therefore identical to Model 3 in Den Haan and De Wind (2012). The optimization problem is given by

$$\max_{(c_t,a_t)_{t=1}^{\infty}} \mathbb{E}_1 \sum_{t=1}^{\infty} \beta^{t-1}\left(\frac{c_t^{1-\gamma}-1}{1-\gamma} - P(a_t)\right),$$

$$\text{s.t.} \quad \begin{aligned} c_t + a_t/(1+r) &= a_{t-1} + e^{z_t}, \\ z_t &= \bar{z} + \varepsilon_t, \\ \varepsilon_t &\sim N(0,\sigma^2), \\ a_0 & \text{ given}, \end{aligned}$$

where c_t stands for the agents consumption, e^{z_t} represents exogenous and random income and r is the exogenous interest rate. The variable a_t denotes the amount of chosen assets in period t, we assume that a_0 is given. The amount of assets is allowed to be negative, so the agent can borrow. The function P is given by

$$P(a_t) = \frac{\eta_1}{\eta_0} e^{-\eta_0 a_t} + \eta_2 a_t.$$

Note that it is decreasing in its argument and thus penalizes utility when the agent decides to borrow. We write $x_t = a_{t-1} + e^{z_t}$ for the amount of cash on hand at time t. Note that this DSGE model has a univariate state equation in x_t, because the z_t are independent.

Our calibration is copied from the original paper and given in Table 1. The value of

r	γ	$\bar{z}$	σ	β	η_0	η_1	η_2
0.03	3	0.4	0.1	0.9	20	0.04464	0.00352

Table 1: The choice of parameter values for the Deaton model.

β is low to make agents impatient and ensure that borrowing constraints have sufficient effect on the decision process. The value of σ is chosen large, because the model describes single agent/household behavior and thus works with idiosyncratic uncertainty. The values of η_1 and η_2 are chosen such that a_t has the same moments as in Deaton Angus (1991). We refer to Den Haan and De Wind (2012) for a more detailed discussion on the model and choice of parameters.

We use a second order perturbation approximation to obtain

$$x_{t+1} - x_{ss} = 0.01 + 0.42(x_t - x_{ss}) + 1.02(x_t - x_{ss})^2 + e^{z_{t+1}}$$

and values for τ given by $\tau^* = 1.08$ and $\hat{\tau} = 0.98$. The innovations in the model are strictly positive, so we cannot use Assumption C to ensure stability of transformed perturbation sample paths. Instead we use Assumption B, which is easy to check in univariate cases. Note that all parameters are positive and the autoregressive parameter is smaller than one. It immediately follows that the transformed perturbation approximation is able to reach any sufficiently large point and thus we have an open interval of reachable points and Assumption B1 is satisfied. All the other Assumptions in A and B are easily checked. Therefore we obtain all the desired stability results from Theorem 3.2.

To compare the approximate policy functions we plot in Figure 3a the expected value of next-period's cash on hand $\mathbb{E}(x_{t+1} \mid x_t)$, because this directly reveals whether the dynamics are stable or not. The true policy function has a single stable fixed point (an attractor). In contrast, the second order perturbation policy function has a second fixed point (a repeller). This second intersection with the $y = x$ line is located above the true steady state. Sample paths produced by the second order perturbation function eventually reach the state space to the right of the repeller, after which they are expected to diverge, and eventually do with probability one. Since the second fixed point is relatively close to the true steady state this also frequently occurs in our finite time simulated paths, making second order perturbation infeasible. The transformed perturbation policy function solves the problem as it negates the second order monomial fast enough to ensure that no second fixed-point is created. The optimal and plug in values for τ, while irrelevant for stability, therefore create a policy function that generates very similar dynamics as the true policy function. Figure 3c displays the same functions as in Figure 3a, but focussed on the relevant part of the state space when using stable methods. In addition we have added a scatter plot of the pruning sample path. From this plot it becomes immediately apparent that pruning does not deliver a policy function on the original state space, as we have different updates for the same starting value. Moreover, it can be seen that pruning on average is less accurate than both the transformed perturbation methods. The policy function corresponding to the optimal τ^* can be seen to be slightly more accurate than the plug in $\hat{\tau}$. This is extra apparent when we look at the pointwise errors between the true path and the perturbation respective transformed perturbation approximations in Figure 3b.

It's not surprising that the resulting transformed perturbation sample paths are very close to the true ones. The sample path accuracy results are summarised in Table 2, where we report maximum and mean absolute path errors in addition to Euler errors. Here we see that second order perturbation explodes, so sample paths created by this

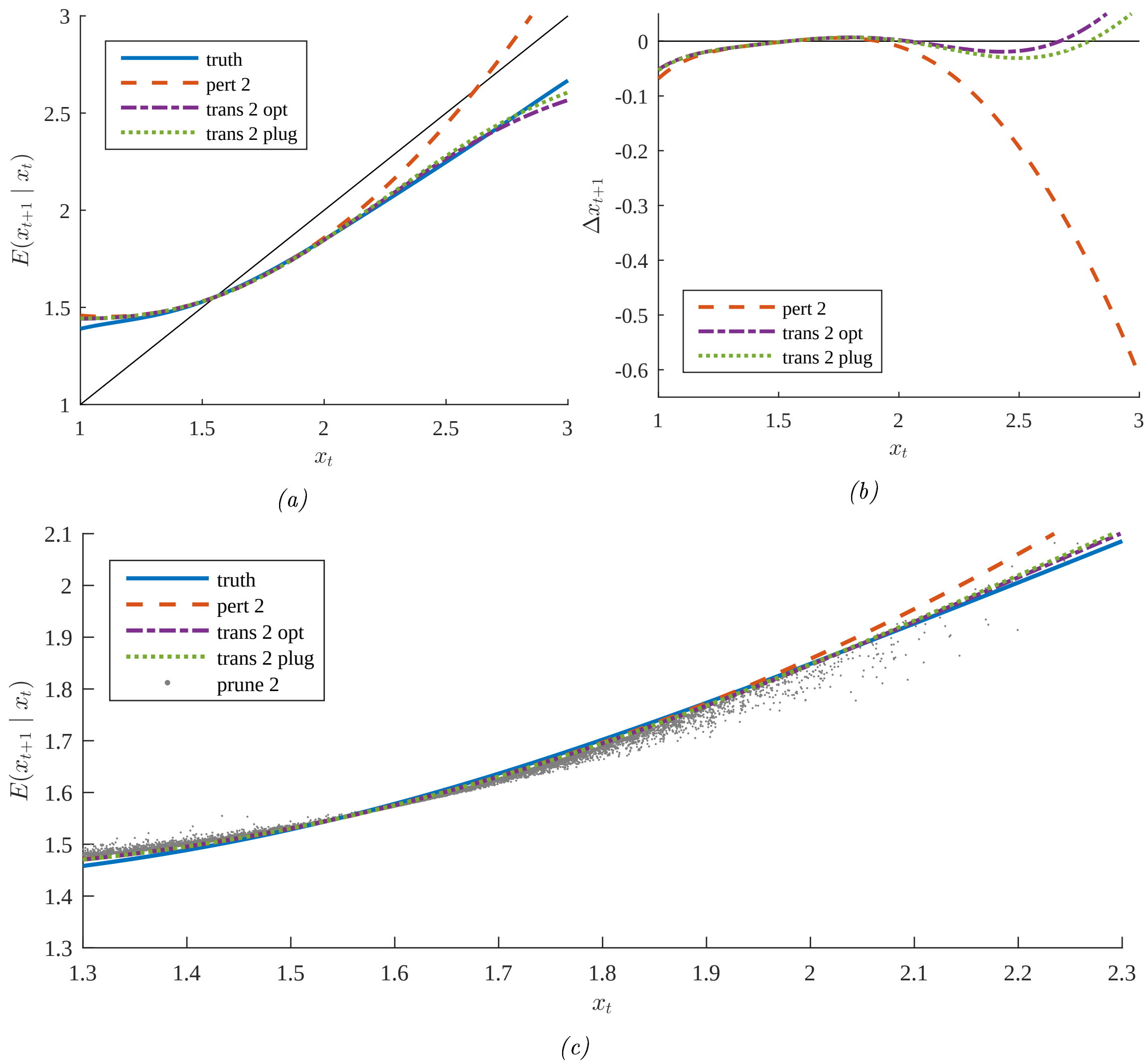


Figure 3: Expected policy functions for x_t in the Deaton model generated by a second order perturbation approximation and the transformed perturbation method for both the optimal τ^* and the plug in $\hat{\tau}$. Figure 3a shows the actual policy functions, Figure 3b shows the pointwise errors with respect to a close approximation of the true policy function and Figure 3c zooms in on the relevant part of the state space to compare the previous methods to pruning.

approximation are unusable. Therefore we need a stable approximation approach. The transformed perturbation approximation performs better than pruning and much better than linear approximation on all criteria. Note that the maximum and mean path errors for the transformed perturbation are about half of those for the pruning approximation, in both the asset and consumption paths.

The difference in accuracy is extra apparent when we look at the cumulative path errors, see Figure 4, which are significantly smaller for our method. This accumulation of inaccuracy then leads to larger errors when we compute some of the sample moments, which can be found in Table 3. Here we see that first order perturbation performs a lot worse than the other methods on the asset moments, which was to be expected, as it missed the nonlinear effects. Transformed perturbation is more accurate than pruning for all moments, especially for ones concerning the assets where we see improvement up to a

	Path errors				Euler errors
	a_t		c_t		
	max	mean	max	mean	
Perturbation 1	132	38.4	10.7	1.06	3.11
Perturbation 2	∞	∞	∞	∞	-
Transformed 2 optimal	**53.0**	6.54	**3.29**	**0.31**	**0.28**
Transformed 2 plug-in	54.4	**6.50**	3.40	**0.31**	0.29
Pruning 2	123	13.6	6.42	0.69	0.51

Table 2: Absolute sample path and Euler errors for the Deaton model. Path errors are compared to a projection approximation and given in percentages. Euler errors are also scaled by 10^2. The results are based on a time path of 10^4 observation with a burn in time of 500 observations.

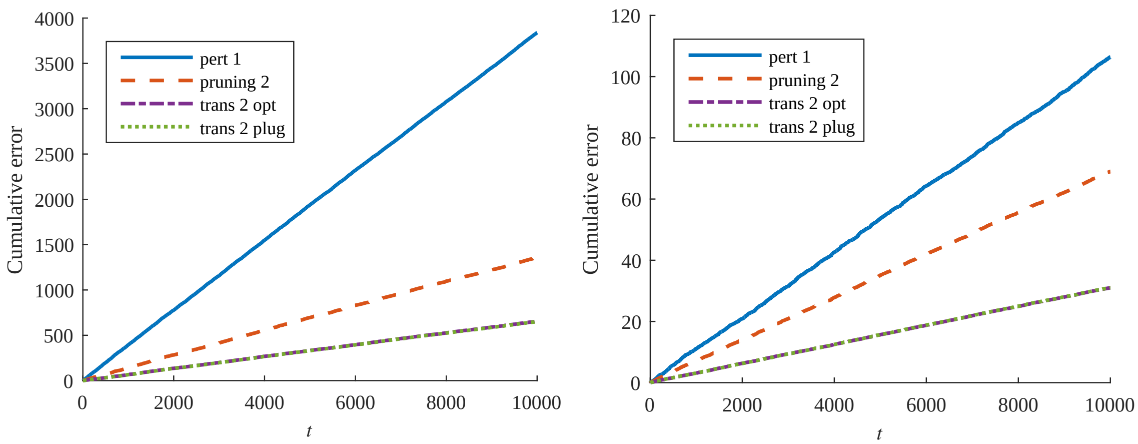


Figure 4: Cumulative paths errors for the number of assets in the left panel and consumption in the right panel. Errors are calculated by a close approximation of the true policy function.

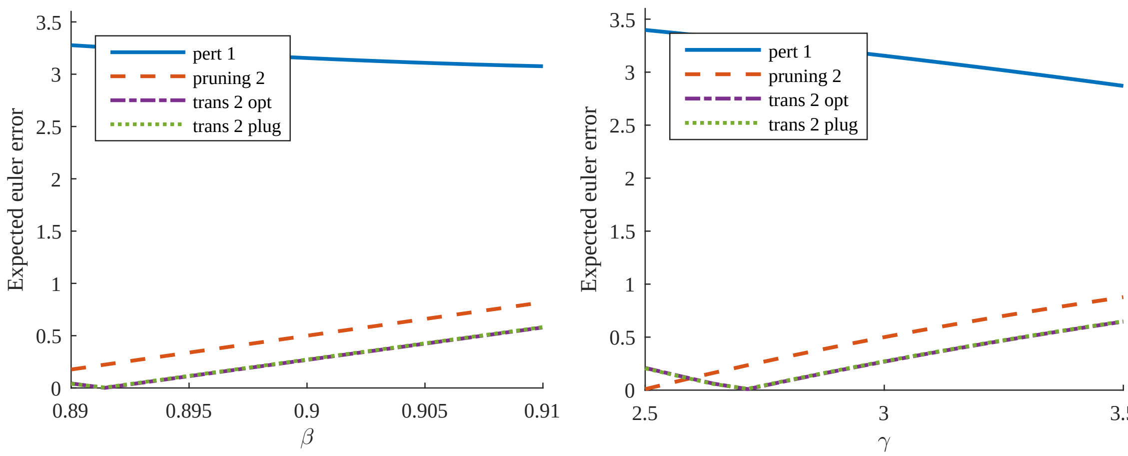
factor ten. Surprising is that the plug-in $\hat{\tau}$ transformed policy function performs better on the moments than the optimal τ^* transformed policy function.

Performance in a parameter estimation scenario

When researchers are interested in estimating parameters, it is important to ensure that the employed approximation method is accurate across a wide range of parameter values. It is thus important to investigate what happens to the accuracy of our approximation methods when we move the parameters away from an initial calibrated parameter value. Figure 5 plots the expected Euler errors for varying values of β and γ. Note that, as described in Section 2.4, for the optimal transformed perturbation method we have kept the initial calculated optimal τ^*, while the plug in transformed perturbation method updates $\hat{\tau}$ along with the parameters. We see in Figure 5 that the expected Euler errors for both the transformed perturbation methods are smaller than those for the pruning method on a significant area around the calibration. This implies that each transformed perturbation method outperforms the pruning method in an estimation setting when the initial parameters have been set sufficiently close to the true ones. The two transformed perturbation methods have such similar Euler errors, because the plug-in$\hat{\tau}$ does not vary much as we change the parameters and stays especially close to the optimal τ^*.

	Sample moments							
	$\mu(a_t)$	$\mu^2(a_t)$	$\mu^3(a_t)$	$\mu^4(a_t)$	$\mu(c_t)$	$\mu^2(c_t)$	$\mu^3(c_t)$	$\mu^4(c_t)$
True	0.083	0.016	0.004	0.001	1.502	2.264	3.423	5.192
Perturbation 1	58.1	62.2	84.2	89.7	0.09	0.11	0.03	**0.16**
Transformed 2 optimal	0.70	7.58	7.83	6.79	**0.00**	**0.03**	**0.09**	0.17
Transformed 2 plug-in	**0.43**	**6.89**	**6.25**	**3.77**	**0.00**	**0.03**	**0.09**	0.17
Pruning 2	5.88	23.50	31.66	38.27	0.01	0.06	0.20	0.40
	Cross moments							
	$\mu(a_tc_t)$	$\mu(a_tc_t^2)$	$\mu(a_tc_t^3)$	$\mu(a_t^2c_t)$	$\mu(a_t^2c_t^2)$	$\mu(a_t^3c_t)$		
True	0.13	0.20	0.31	0.03	0.04	0.006		
Perturbation 1	57.3	56.5	55.7	63.0	63.6	84.1		
Transformed 2 optimal	0.96	1.18	1.35	7.45	7.30	7.65		
Transformed 2 plug-in	**0.69**	**0.90**	**1.07**	**6.73**	**6.57**	**6.03**		
Pruning 2	6.30	6.58	6.75	23.2	22.8	31.1		

Table 3: Sample and cross moments up to fourth order for the Deaton model. The true row presents the moments given by a close approximation. The other moments are given as absolute percentage differences from the true ones. The results are based on a time path of 10^4 observation with a burn in time of 500 observations.



(a) *(b)*

Figure 5: Expected Euler errors for the Deaton model on an area around the calibrated parameter values. Figure 5a portrays the results when changing β and Figure 5b when changing γ.

5.2.2 The Matching model

The second model we examine is a matching model also featured in Den Haan and De Wind (2012). The model has two types of agents, workers and entrepreneurs, both of which are members of the same representative household. The household earns wages and firm profits from its members at the end of each period. These are then distributed among the members for consumption.

Firms: The main decision is made by a representative entrepreneur who tries to maximise future discounted firm profits. The maximisation problem is given by

$$
\begin{aligned}
&\max_{(n_t,v_t)_{t=1}^{\infty}} \mathbb{E}_1 \sum_{t=1}^{\infty} \beta^{t-1} \left(\frac{c_t}{c_1}\right)^{-\gamma} \left(\left(e^{z_t} - w\right) n_{t-1} - \psi v_t\right), \\
&\text{s.t.} \qquad n_t = (1-\rho_n) n_{t-1} + p_{f,t} v_t, \\
&\qquad\qquad z_t = \begin{cases} z_{t-1} & \text{with probability } \rho_z \\ -z_{t-1} & \text{with probability } 1-\rho_z \end{cases}, \\
&\qquad\qquad n_0, z_1 \text{ given.}
\end{aligned}
$$

Here c_t is the consumption level of the household, n_t is the number of employees at the end of period t, v_t is the number of vacancies set by the firm, $p_{f,t}$ is the number of matches per vacancy, w is the wage rate, ψ is the cost of placing a vacancy and ρ_n is the exogenous separation rate. Each worker produces e^{z_t}, which means that the profit per worker is given by $e^{z_t} - w$. The random variable z_t can only take on two values, which we denote $-\zeta$ and $+\zeta$. This is an artificial simplification introduced in Den Haan and De Wind (2012) enabling us to easily analyse the approximation methods to the model in a graphical manner. Alternatively, one can use a standard autoregressive updating function for z_t. Finally, the firm takes the number of matches $p_{f,t}$ as given.

Consumers: The household consumes the whole income earned by its members. That is,

$$
c_t = w n_{t-1} + \left(e^{z_t} - w\right) n_{t-1} - \psi v_t = e^{z_t} n_{t-1} - \psi v_t.
$$

Matching market: The number of hires per vacancy is determined on a matching market where the firms and $1 - n_{t-1}$ unemployed workers search for a match. The total number of matches is given by

$$
m_t = \phi_0 (1 - n_{t-1})^{\phi} v_t^{1-\phi}.
$$

This implies that the total number of matches per vacancy is given by

$$
p_{f,t} = \frac{m_t}{v_t} = \phi_0 \left(\frac{1 - n_{t-1}}{v_t}\right)^{\phi}.
$$

The model requires some restrictions on the parameters to ensure that a solution in the interior of the domain exists and thus that the policy function is smooth. Our choice of parameter values is again taken from Den Haan and De Wind (2012) and given in Table 4. See the original paper for a detailed discussion on the matching model, the parameter values and further references.

γ	w	ψ	ρ_n	ρ_z	ζ	σ	β	ϕ_0	ϕ
4.5	0.973	0.4026	0.0368	0.975	0.0224	0.007	0.99	0.7	0.5

Table 4: The choice of parameter values for the Matching model.

A second order perturbation approximation of the state equation delivers

$$
\begin{aligned}
n_{t+1} - n_{ss} = {} & 0.95 + 0.46(n_t - n_{ss}) + 0.52 z_{t+1} \\
& - 2.92(n_t - n_{ss})^2 - 6.57(n_t - n_{ss}) z_{t+1} - 1.01 z_{t+1}^2
\end{aligned}
$$

and we find

$$\tau^* = 26.1 \qquad \text{and} \qquad \hat{\tau} = 13.6.$$

The updating equation for the exogenous state variable z_t is not of the type (2). One can extend the theory in a rather straightforward way to also apply to general Markov chain updating equations for the exogenous state variables, but we chose not to do this to keep the assumptions and proofs relatively clear and concise. Note that if we would have chosen a standard autoregressive process of order one for $(z_t)_{t\geq 0}$, then Assumptions A and C can easily seen to be satisfied as we have a univariate system. Therefore, in that case, we would have obtained all the desired stability results from Theorem 3.2.

The control variables can be explicitly calculated once the path for the single state variable, the number of employees, is known. We therefore compare the approximation methods according to their best performance: either calculating the control variables directly, or approximating the observation equation. We compare the transformed perturbation and regular perturbation approximation in Figure 6. Figure 6a shows the policy functions for the number of employees in the two possible scenarios for z_t. The case $z_t = -\zeta$ is the crucial one here, as the regular perturbation approximation stays below the $y = x$ line and therefore does not intersect it. This implies that the second order perturbation sample paths for n_t tend to minus infinity if z_t is equal to $-\zeta$ for many consecutive times. The case $z_t = +\zeta$ goes to minus infinity for values of n_t much smaller than portrayed in the figure. Hence, once n_t has become small enough it has no chance of recovering and thus sample paths diverge to minus infinity with probability one. As in the previous example, the explosive behaviour is encountered in our finite time sample paths with a high frequency, rendering regular perturbation infeasible. The transformed perturbation policy function avoids the problem described for both values of τ as they both scale down the second order monomial fast enough to ensure that the policy functions cross the $y = x$ line at a unique point, like the true policy function. The dynamics of our approximated systems therefore closely mimic the true dynamics for n_t. Figure 6c again zooms in on the relevant part of the state space when using the stable solution methods and includes a scatter plot of the pruning sample path. Again, we are reminded that pruning does not provide a policy function on the original state space. Moreover, pruning provides less accurate updates, especially for large value of n_t in the case $z_t = +\zeta$ and small values in the case $z_t = -\zeta$. The policy function corresponding to the optimal τ^* is clearly the most accurate method in our comparison, which is extra clear when we look at the pointwise errors between the true path and the perturbation respective transformed perturbation approximations in Figure 6b.

The graphical results are strengthened by studying the sample path errors in Table 5. Here we see that the transformed perturbation approximation is both in extreme cases and on average performing better than both perturbation and pruning. The improvement compared to perturbation is not surprising given the nonlinearity of the plots in Figure 6. This time the optimal transformed perturbation method performs better than the plug in approximation. It is also more than a factor ten times better on average than pruning for the number of employees and more than a factor three times better on average than pruning on consumption paths.

We emphasize the gravity of the difference in accuracy by plotting the cumulative path errors in Figure 7. This total difference in accuracy then again leads to a large difference in sample moment accuracy, which is summarised in Table 6. Like before we see that the transformed perturbation method, especially the optimal one, is best at mimicking the

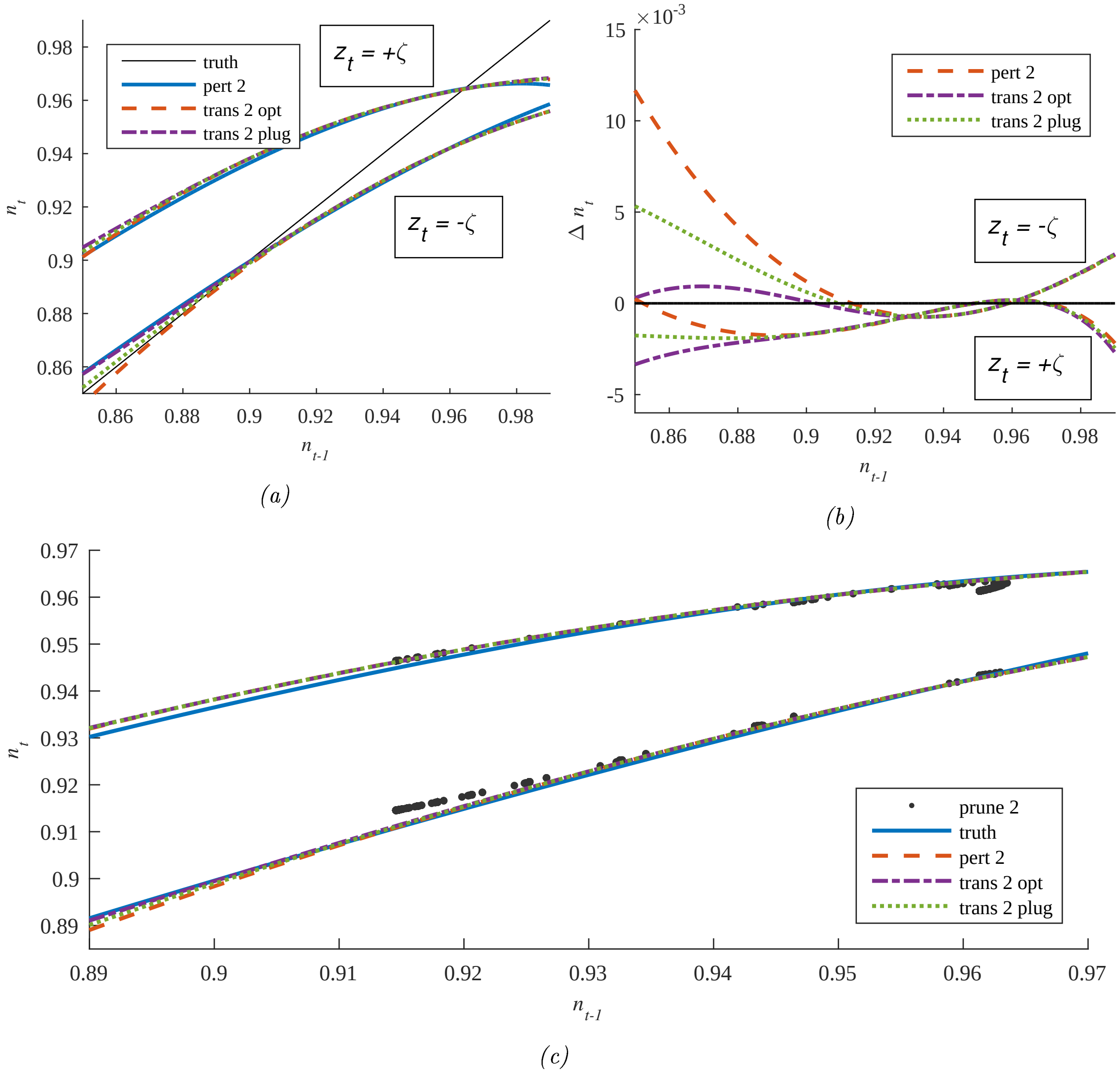


Figure 6: Policy functions for n_t in the matching model generated by a second order perturbation approximation and the transformed perturbation method. Figure 6a shows the actual policy functions, Figure 6b shows the pointwise errors with respect to a close approximation of the true policy function and Figure 3c zooms in on the relevant part of the state space to compare the previous methods to pruning.

dynamics of the sample paths. Note that both the optimal and transformed perturbation method outperform pruning on all moments, especially for the higher order moments, where pruning loses relatively more accuracy by ignoring higher order effects. Optimal transformed perturbation outperforms pruning up to a factor forty for the fourth order moments of consumption, while plug-in transformed perturbation outperforms pruning by a factor three for most moments.

Performance in a parameter estimation scenario

Once more we investigate the accuracy of the discussed methods in an area around the calibrated parameter values. Figure 8 plots the expected Euler errors for varying values of

	Path errors				Euler errors
	n_t		c_t		
	max	mean	max	mean	
Perturbation 1	3.20	1.89	3.53	1.80	0.26
Perturbation 2	∞	∞	∞	∞	-
Transformed 2 optimal	**0.26**	**0.07**	**0.64**	0.32	0.08
Transformed 2 plug-in	0.71	0.25	0.97	**0.23**	**0.04**
Pruning 2	1.79	0.95	1.76	0.95	0.10

Table 5: Absolute sample path and Euler errors for the matching model. Path errors are compared to a close approximation of the truth and given in percentages. Euler errors are also scaled by 10^2*. The results are based on a time path of* 10^4 *observation with a burn in time of* 500 *observations.*

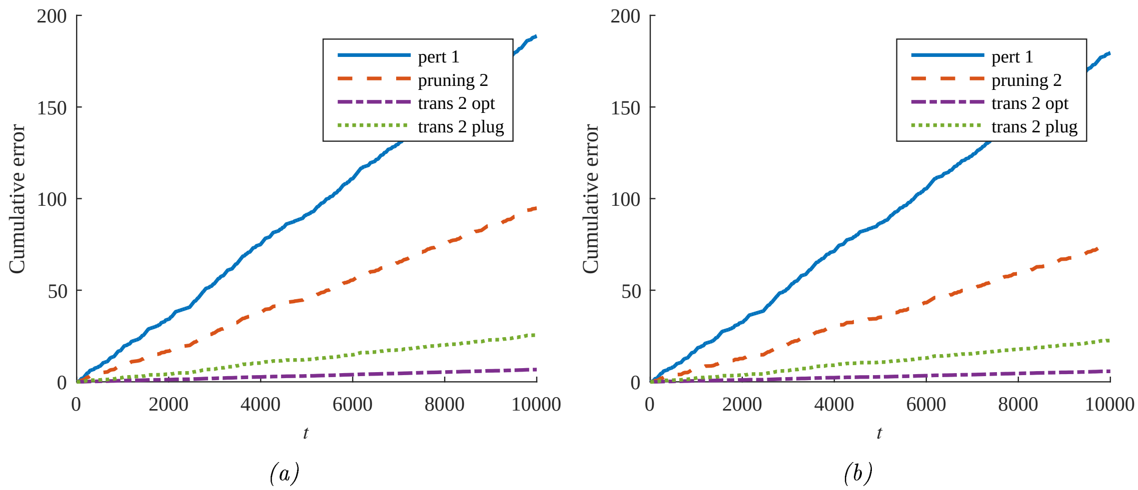


Figure 7: Cumulative paths errors for the number of employees in the left panel and consumption in the right panel. Errors are calculated by a close approximation of the true policy function.

β and γ while keeping the steady state values for the number of employees, the number of matches per unemployed worker and the number of matches per vacancy equal. As in the previous example we fix the optimal τ^* at the initial derived value at the calibrated parameters, while the plug in $\hat{\tau}$ is updated along with the parameters. Figure 8 shows us that the expected Euler errors for each transformed perturbation method is smaller than those for the pruning method on a relevant area around the calibration. Therefore, we again conclude that an estimation procedure using the transformed perturbation method improves accuracy over using either linear perturbation or pruning when the starting values are decently close to the true parameters.

	Sample moments							
	$\mu(n_t)$	$\mu^2(n_t)$	$\mu^3(n_t)$	$\mu^4(n_t)$	$\mu(c_t)$	$\mu^2(c_t)$	$\mu^3(c_t)$	$\mu^4(c_t)$
True	0.93	0.87	0.81	0.76	0.91	0.83	0.76	0.70
Perturbation 1	1.90	3.77	5.62	7.43	1.81	3.57	5.27	6.91
Transformed 2 optimal	**0.03**	**0.05**	**0.07**	**0.10**	**0.02**	**0.04**	**0.05**	**0.06**
Transformed 2 plug-in	0.23	0.45	0.65	0.83	0.20	0.39	0.55	0.70
Pruning 2	0.68	1.30	1.88	2.40	0.65	1.23	1.73	2.16
	Cross moments							
	$\mu(n_tc_t)$	$\mu(n_tc_t^2)$	$\mu(n_tc_t^3)$	$\mu(n_t^2c_t)$	$\mu(n_t^2c_t^2)$	$\mu(n_t^3c_t)$		
True	0.85	0.78	0.71	0.79	0.73	0.74		
Perturbation 1	3.67	5.39	7.04	5.50	7.17	7.30		
Transformed 2 optimal	**0.04**	**0.06**	**0.07**	**0.07**	**0.08**	**0.09**		
Transformed 2 plug-in	0.42	0.59	0.73	0.62	0.77	0.80		
Pruning 2	1.27	1.78	2.22	1.83	2.28	2.34		

Table 6: Sample and cross moments up to fourth order for the matching model. The true row presents the moments given by a close approximation. The other moments are given as absolute percentage differences from the true ones. The results are based on a time path of 10^4 *observation with a burn in time of* 500 *observations.*

A Appendix

A.1 Proofs of Section 3

We study the transformed perturbation method, as indicated in Section 3, as its asymptotic linear process plus a deviation (9). The deviation is bounded in $\mathbf{x}$ as $\Phi_\tau(\tilde{\mathbf{x}})$ dominates the function far away from the origin. The following result gives a uniform upper bound to the size of the deviation over $\mathcal{X}$.

Proposition A.1. *There exists a constant $c \geq 0$ that does not depend on τ, such that*

$$\sup_{\mathbf{x}\in\mathcal{X}} \|D(\mathbf{x}, z)\| \leq c \sum_{j=0}^{m} \tau^{-j/2} \left(\sum_{i=0}^{m-j} \|z\|^i \right).$$

PROOF. In this proof we specifically choose $\|\cdot\|$ equal to the Euclidean matrix norm $\|\cdot\|_e$. This matrix norm is a crossnorm, i.e. it is multiplicative on Kronecker products, see for example Lancaster and Farahat (1972). This implies, together with sub-additivity and sub-multiplicativity, that

$$\begin{aligned}
\|D(\mathbf{x}, \mathbf{z})\| &\leq \sum_{i=2}^{m} \|H_i\| \|\mathbf{v}\|^i \Phi_\tau(\tilde{\mathbf{x}}) \\
&\leq \left(\max_{2\leq i\leq m} \|H_i\| \right) \sum_{i=2}^{m} \sum_{j=0}^{i} \|\mathbf{x}\|^j \|\mathbf{z}\|^{i-j} \Phi_\tau(\tilde{\mathbf{x}}) \\
&\leq \left(\max_{2\leq i\leq m} \|H_i\| \right) \sum_{j=0}^{m} \|\mathbf{x}\|^j \Phi_\tau(\tilde{\mathbf{x}}) \left(\sum_{i=0}^{m-j} \|\mathbf{z}\|^i \right).
\end{aligned}$$

Next, note that

$$\|\mathbf{x}\|^j \Phi_\tau(\tilde{\mathbf{x}}) \leq \|\mathbf{x}\|^j e^{-\tau \|\mathbf{x}\|_e^2 / \max\{\mathbf{x}_{ss}\}},$$

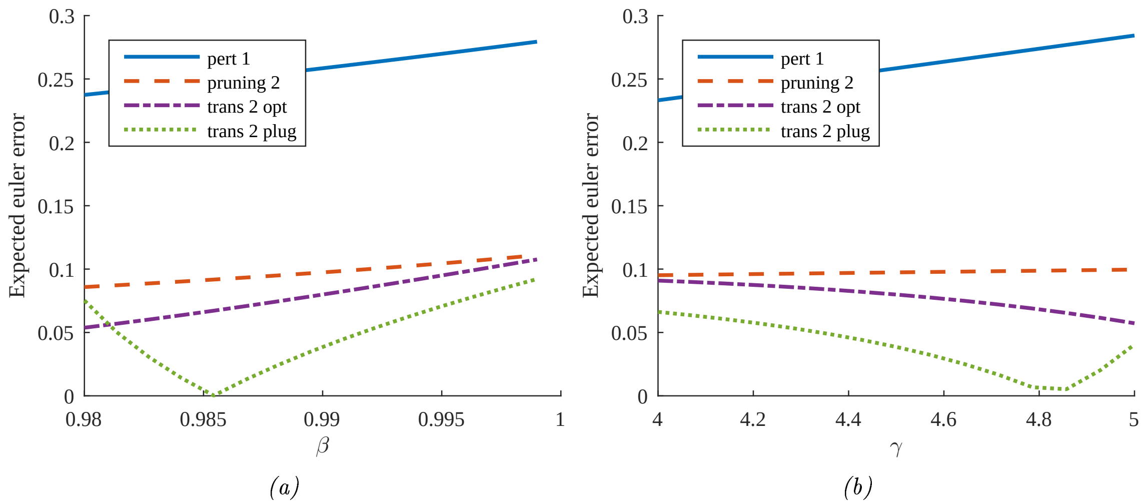


Figure 8: Expected Euler errors for the matching model on an area around the calibrated parameter values. Figure 8a portrays the results when changing β and Figure 8b when changing γ while keeping the steady state values for the number of employees, the number of matches per unemployed worker and the number of matches per vacancy equal.

which is a univariate function in $\|\mathbf{x}\|_e$, since we chose $\|\cdot\|$ equal to $\|\cdot\|_e$. It is straightforward to verify that this function is maximised at $\|\mathbf{x}\|_e^2 = \frac{j \max\{\mathbf{x}_{ss}\}}{2\tau}$ and thus there exists a constant $\tilde{c}$ that does not depend on τ or $\mathbf{x}$ such that

$$\sup_{\mathbf{x}\in\mathcal{X}} \|\mathbf{x}\|^j \Phi_\tau(\tilde{\mathbf{x}}) \leq \tilde{c}\tau^{-j/2} \qquad \text{for all } 0 \leq j \leq m.$$

■

A.1.1 Proof of Theorem 3.1

Assumptions A1 and A3 imply by Theorem 3.1 in Bougerol (1993) and the monotone convergence theorem that there exists a unique stationary ergodic solution $(\mathbf{z}_t^*)_{t\in\mathbb{N}}$ to (2) with $\mathbb{E}\|\mathbf{z}_t^*\|^{rm} < \infty$. Moreover, $\|\mathbf{z}_t - \mathbf{z}_t^*\|$ converges exponentially almost surely to zero as $t \to \infty$, which implies that

$$\liminf_{t\to\infty} \|\mathbf{z}_t\| < \infty \qquad \text{a.s.}$$

and that, for every realisation, there exists a constant $d > 0$ such that $\|\mathbf{z}_t\|^i \leq \|\mathbf{z}_t^*\|^i + d$ for all $t \geq 0$ and $0 \leq i \leq m$.

Next, we repeatedly expand the term $H_\mathbf{x}\mathbf{x}$ in (7) to obtain the following expression for the transformed perturbation path:

$$\mathbf{x}_t = H_\mathbf{x}^t \mathbf{x}_0 + \sum_{k=0}^{t-1} H_\mathbf{x}^k \left(H_0 + H_\mathbf{z}\mathbf{z}_{t-k} + D(\mathbf{x}_{t-1-k}, \mathbf{z}_{t-k})\right).$$

We now use Proposition A.1 to bound the deviation terms and then use the bounds on the path $(\mathbf{z}_t)_{t\geq 0}$ to obtain

$$\begin{aligned}
&\|\mathbf{x}_t\| - \|H_{\mathbf{x}}\|^t\|\mathbf{x}_0\| \\
&\quad \leq \sum_{k=0}^{t-1} \|H_{\mathbf{x}}\|^k \left(\|H_0\| + \|H_{\mathbf{z}}\|\|\mathbf{z}_{t-k}\| + c\sum_{j=0}^{m} \tau^{-j/2} \left(\sum_{i=0}^{m-j} \|\mathbf{z}_{t-k}\|^i \right) \right) \\
&\quad \leq \sum_{k=0}^{t-1} \|H_{\mathbf{x}}\|^k \left(\|H_0\| + \|H_{\mathbf{z}}\|(\|\mathbf{z}^*_{t-k}\| + d) + c\sum_{j=0}^{m} \tau^{-j/2} \left(\sum_{i=0}^{m-j} \|\mathbf{z}^*_{t-k}\|^i + d \right) \right) \qquad (16)
\end{aligned}$$

Next we artificially extend $(\mathbf{z}^*_t)_{t\geq 0}$ to a stationary ergodic sequence $(\mathbf{z}^*_t)_{t\in\mathbb{Z}}$ and then note that (16) is bounded by

$$Y_t := \sum_{k=0}^{\infty} \|H_{\mathbf{x}}\|^k \left(\|H_0\| + \|H_{\mathbf{z}}\|(\|\mathbf{z}^*_{t-k}\| + d) + c\sum_{j=0}^{m} \tau^{-j/2} \left(\sum_{i=0}^{m-j} \|\mathbf{z}^*_{t-k}\|^i + d \right) \right).$$

The term within the brackets is stationary ergodic by Krengel's lemma, see Proposition 4.3 in Krengel (1985), and the fact that $(\mathbf{z}^*_t)_{t\in\mathbb{Z}}$ is stationary ergodic. Moreover it has a finite log moment since $\mathbb{E}\|\mathbf{z}^*_t\|^{rm} < \infty$. Next, we can choose a matrix norm such that $\|H_{\mathbf{x}}\| < 1$ by Assumption A2. Therefore, the infinite sum converges almost surely by Proposition 2.5.1 of Straumann (2005). Again, the sequence $(Y_t)_{t\in\mathbb{Z}}$ is stationary ergodic by Krengel's lemma and thus there almost surely exists an $M > 0$ such that $\{Y_t \leq M\}$ occurs for infinitely many $t > 0$. We conclude that

$$\liminf_{t\to\infty} ||\mathbf{x}_t|| \leq M < \infty.$$

A.1.2 Proof of Theorem 3.2

We study the processes $(\mathbf{z}_t)_{t\geq 0}$ and $(\mathbf{x}_t)_{t\geq 0}$ as a joint Markov process. This section will make extensive use of Meyn and Tweedie (1993). We will first assume that $(\mathbf{z}_t, \mathbf{x}_t)_{t\geq 0}$ is a ψ-irreducible and aperiodic T-chain. See sections 4.2, 5.4 and 6.2 of Meyn and Tweedie (1993) for a detailed discussion on these properties.

Proposition A.2. *Suppose $(\mathbf{z}_t, \mathbf{x}_t)_{t\geq 0}$ is a ψ-irreducible and aperiodic T-chain and let Assumptions A and B hold. Then all the results of Theorem 3.2 hold.*

Proof. We will check the drift condition for t-step transitions, which is described in condition *(iii)* of Theorem 1 in Saïdi and Zakoian (2006), adapted from Theorem 19.1.3 in Meyn and Tweedie (1993) and originally suggested by Tjøstheim (1990). The condition states that we need to find a non-negative function $V : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ and a $t \in \mathbb{N}$ such that

$$\frac{\mathbb{E}\left(V(\mathbf{x}_t, \mathbf{z}_t) \mid \mathbf{x}_0 = \mathbf{x}, \mathbf{z}_0 = \mathbf{z}\right)}{V(\mathbf{x}, \mathbf{z})} \qquad (17)$$

is finite on a compact set $C \subseteq \mathcal{X} \times \mathcal{Z}$ and smaller than one outside of C. Note that the set C actually has to be petite, but all compact sets are petite in a ψ-irreducible T-chain, Theorem 6.2.5 in Meyn and Tweedie (1993). It then follows by Theorem 1 in Saïdi and Zakoian (2006) that there exists a unique stationary ergodic solution $(\mathbf{x}^*_t, \mathbf{z}^*_t)_{t\geq 0}$ that is geometrically ergodic and has the required moments, given our choice for V. Absolute regularity follows from Theorem 1 in Davydov (1974) and the laws of large numbers follow from Theorem 17.0.1 in Meyn and Tweedie (1993). The reason that we resort to t-step, instead of 1-step, transitions is that Assumption A1 and Assumption A2 do not guarantee that there exists a matrix norm such that both $\|\mathbf{\Lambda}\| < 1$ and $\|H_{\mathbf{x}}\| < 1$. Assumption A1 can ensure that there exists a matrix norm such that $\|\mathbf{\Lambda}\| < 1$, but then Assumption A2 only provides the existence of a $t \in \mathbb{N}$ such that $\|H_{\mathbf{x}}^t\| < 1$ by Gelfand's formula.

We adopt the ideas of Cline and Pu (1999) and use the test function

$$V(\mathbf{x}, \mathbf{z}) = 1 + (\|\mathbf{x}\| + \omega\|\mathbf{z}\|^m)^r,$$

where we will choose $\omega > 0$ sufficiently large. If $r \leq 1$, then $(\|\mathbf{x}\| + \omega\|\mathbf{z}\|^m)^r \leq \|\mathbf{x}\|^r + \omega^r\|\mathbf{z}\|^{rm}$. We prove the theorem for the case $r \geq 1$, as it is the harder case. In that case Minkowski's inequality provides the upper bound

$$\mathbb{E}((\|\mathbf{x}_t\| + \omega\|\mathbf{z}_t\|^m)^r \mid \mathbf{x}_0, \mathbf{z}_0) \leq \left(\mathbb{E}(\|\mathbf{x}_t\|^r \mid \mathbf{x}_0, \mathbf{z}_0)^{\frac{1}{r}} + \omega\mathbb{E}(\|\mathbf{z}_t\|^{rm} \mid \mathbf{x}_0, \mathbf{z}_0)^{\frac{1}{r}}\right)^r.$$

We start by bounding the second expectation. Note that the expectations $\mathbb{E}\|\boldsymbol{\sigma\eta\varepsilon}_s\|^{rm}$ are bounded for all $s \in \{1, \dots, t\}$ by Assumption A3. Expanding backwards and working out brackets then gives

$$\begin{aligned} \mathbb{E}\left(\|\mathbf{z}_t\|^{rm} \mid \mathbf{z}_0\right) &\leq \mathbb{E}\left(\left(\|\mathbf{\Lambda}^t\mathbf{z}_0\| + \|\mathbf{\Lambda}^{t-1}\boldsymbol{\varepsilon}_1\| + \cdots \|\boldsymbol{\varepsilon}_t\|\right)^{rm} \mid \mathbf{z}_0\right) \\ &\leq \|\mathbf{\Lambda}\|^{trm}\|\mathbf{z}_0\|^{rm} + o\left(\|\mathbf{z}_0\|^{rm}\right) \qquad \text{as } \|\mathbf{z}_0\| \to \infty. \end{aligned} \tag{18}$$

Next, by Proposition A.1 there exist constants $c_1, c_2 > 0$ such that $\|D(\mathbf{x}_{s-1}, \mathbf{z}_s)\| < c_1 + c_2(1 + \|\mathbf{z}_s\|^m)$ for all $s \in \{1, \dots, t\}$. It then follows again by backwards expansion and the fact that $\|\mathbf{z}_s\| \leq 1 + \|\mathbf{z}_s\|^m$ that there exist constants $d_1, d_2 > 0$ such that

$$\begin{aligned} \mathbb{E}\left(\|\mathbf{x}_t\|^r \mid \mathbf{x}_0, \mathbf{z}_0\right) &\leq \mathbb{E}\left(\left\|H_{\mathbf{x}}^t\mathbf{x}_0 + \sum_{k=0}^{t-1} H_{\mathbf{x}}^k\left(H_0 + H_{\mathbf{z}}\mathbf{z}_{t-k} + D(\mathbf{x}_{t-1-k}, \mathbf{z}_{t-k})\right)\right\|^r \mid \mathbf{x}_0, \mathbf{z}_0\right) \\ &\leq \mathbb{E}\left(\left(\|H_{\mathbf{x}}^t\|\|\mathbf{x}_0\| + d_1 + d_2\sum_{k=0}^{t-1}\|\mathbf{z}_{t-k}\|^m\right)^r \mid \mathbf{x}_0, \mathbf{z}_0\right) \\ &\leq \left(\|H_{\mathbf{x}}^t\|\|\mathbf{x}_0\| + O\left(\|\mathbf{z}_0\|^m\right)\right)^r \qquad \text{as } \|\mathbf{z}_0\| \to \infty. \end{aligned}$$

The last inequality follows by repeated application of Minkowski's inequality in combination with the same calculations as in (18). Filling everything in then upper bounds (17) by

$$\frac{1 + \left(\|H_{\mathbf{x}}^t\|\|\mathbf{x}\| + (\|\mathbf{\Lambda}\|^{tm} + \omega^{-1}O\left(1\right))\omega\|\mathbf{z}\|^m + o\left(\|\mathbf{z}\|^m\right)\right)^r}{1 + (\|\mathbf{x}\| + \omega\|\mathbf{z}\|^m)^r} \qquad \text{as } \|\mathbf{z}\| \to \infty.$$

Recall that $\|H_{\mathbf{x}}^t\| < 1$ and $\|\mathbf{\Lambda}\| < 1$ and choose ω large enough such that $\|\mathbf{\Lambda}\|^{tm} + \omega^{-1}O(1) < 1$ as $\|\mathbf{z}\| \to \infty$. Then we can make the fraction smaller than one if we choose $\|\mathbf{x}\|, \|\mathbf{z}\| > M$ for a sufficiently large M. Let $C = \{(\mathbf{x}, \mathbf{z}) \in \mathcal{X} \times \mathcal{Z} \mid \|\mathbf{x}\|, \|\mathbf{z}\| \leq M\}$, then (17) is bounded over C and smaller than one outside of C. ∎

It remains to be proven that $(\mathbf{z}_t, \mathbf{x}_t)_{t\geq 0}$ is a ψ-irreducible and aperiodic T-chain, which follows from the results of sections 6.0 - 1 of Meyn and Tweedie (1993). We have, similarly to Proposition 6.1.2 and 6.1.3, that Assumption B2 ensures that the Markov chain is strong Feller. It then follows by Proposition 6.1.5 and Assumption B1 that the Markov chain is ψ-irreducible. Finally, we conclude that $(\mathbf{x}_t)_{t\geq 0}$ is an aperiodic T-chain by Lemma 6.1.4 and part (iii) of Theorem 6.0.1.

A.1.3 Proof of Proposition 3.3

It is clear that Assumption C3 implies Assumption B2, so it remains to prove Assumption C also implies Assumption B1. We will prove a stronger statement: Fix any $\mathbf{x}^* \in \mathcal{X}$ then that point is reachable. Let t be the smallest integer such that assumption C1 holds. The approach will be to show that we can find values for $\mathbf{z}_1, \ldots, \mathbf{z}_t$ that bring $\mathbf{x}_t$ arbitrarily close to $\mathbf{x}^*$. It then follows by Assumption C3 that we have positive probability of $\mathbf{x}_t$ being arbitrarily close to $\mathbf{x}^*$.

To find the values for the exogenous state variables, we start by expanding $\mathbf{x}_t$ back in time as

$$\begin{aligned}\mathbf{x}_t = &\sum_{k=0}^{t-1} H_{\mathbf{x}}^k H_0 + H_{\mathbf{x}}^t \mathbf{x}_0 \\ &+ \begin{bmatrix} H_{\mathbf{x}}^{t-1} H_{\mathbf{z}} & \cdots & H_{\mathbf{x}} H_{\mathbf{z}} & H_{\mathbf{z}} \end{bmatrix} \begin{bmatrix} \mathbf{z}_1' & \mathbf{z}_2' & \cdots & \mathbf{z}_t' \end{bmatrix}' \\ &+ \sum_{k=0}^{t-1} H_{\mathbf{x}}^k D(\mathbf{x}_{t-1-k}, \mathbf{z}_{t-k}).\end{aligned}$$

Assumption C1 ensures that we can select n_x linearly independent columns from the matrix $\begin{bmatrix} H_{\mathbf{x}}^{t-1} H_{\mathbf{z}} & \cdots & H_{\mathbf{x}} H_{\mathbf{z}} & H_{\mathbf{z}} \end{bmatrix}$, which we denote $\mathbf{a}_1, \ldots, \mathbf{a}_{n_x}$. Let $A = \begin{bmatrix} \mathbf{a}_1 & \ldots & \mathbf{a}_{n_x} \end{bmatrix}$ and let $\boldsymbol{\delta} = \begin{bmatrix} \delta_1 & \ldots & \delta_{n_x} \end{bmatrix}'$ be the vector consisting of the univariate stochastic variables inside $\begin{bmatrix} \mathbf{z}_1' & \mathbf{z}_2' & \cdots & \mathbf{z}_t' \end{bmatrix}'$ that correspond to the columns $\mathbf{a}_1, \ldots, \mathbf{a}_{n_x}$. Then, by setting the random variables corresponding to the other columns equal to zero, we get

$$\mathbf{x}_t = \sum_{k=0}^{t-1} H_{\mathbf{x}}^k H_0 + H_{\mathbf{x}}^t \mathbf{x}_0 + A\boldsymbol{\delta} + \sum_{k=0}^{t-1} H_{\mathbf{x}}^k D(\mathbf{x}_{t-1-k}, \mathbf{z}_{t-k}). \tag{19}$$

Suppose all the deviations are zero, then we immediately obtain that we need to choose

$$\boldsymbol{\delta} = A^{-1} \left(\mathbf{x}^* - \sum_{k=0}^{t-1} H_{\mathbf{x}}^k H_0 - H_{\mathbf{x}}^t \mathbf{x}_0 \right). \tag{20}$$

Generally, the deviations are nonzero, so that the choice (20) does not guarantee that $\mathbf{x}_t$ is close to $\mathbf{x}^*$. In fact we would obtain

$$\mathbf{x}_t = \mathbf{x}^* + \sum_{k=0}^{t-1} H_{\mathbf{x}}^k D(\mathbf{x}_{t-1-k}, \mathbf{z}_{t-k}). \tag{21}$$

The idea is then as follows. We show that sample paths can reach arbitrarily large values, and then take such a large value to be our starting point $\mathbf{x}_0$. We then show that as the starting point gets larger our choice for $\boldsymbol{\delta}$ will get larger according to (20) and the whole path from $\mathbf{x}_0$ to $\mathbf{x}_t$ will be arbitrarily large. Since deviations converge to zero away from the steady state we conclude that we can get $\mathbf{x}_t$ arbitrarily close to $\mathbf{x}^*$.

Formally, the deviations in (19) are nonlinear, which together with Assumption C2 and the fact that A is invertible means that we can for any starting point $\mathbf{x}_0$ reach a point $\mathbf{x}_t \in \mathcal{X}$ such that $H_{\mathbf{x}}^t \mathbf{x}_t = \sum_{i=1}^{n_x} \lambda_i \mathbf{a}_i$ has all $\lambda_i \in \mathbb{R}$ arbitrarily large. Therefore we can assume the same for our starting point $\mathbf{x}_0$, that is, for all $d > 0$ we can choose $\mathbf{x}_0$ such that $H_{\mathbf{x}}^t \mathbf{x}_0 = \sum_{i=1}^{n_x} \lambda_i \mathbf{a}_i$ with $|\lambda_i| > d$ for all $1 \leq i \leq n_x$. It immediately follows from (20) that each $|\delta_i|$ goes to infinity linearly in d as we increase d.

Next, we show that increasing d ensures that each $\|\mathbf{x}_{t-j}\|$ for $0 < j < t$ becomes arbitrarily large. Let $A^{(j)}$ and $\boldsymbol{\delta}^{(j)}$ be the sub-matrix respective sub-vector of A and $\boldsymbol{\delta}$ such that for partially expanding $\mathbf{x}_t$ we have

$$\mathbf{x}_t = \sum_{k=0}^{j-1} H_{\mathbf{x}}^k H_0 + H_{\mathbf{x}}^j \mathbf{x}_{t-j} + A^{(j)}\boldsymbol{\delta}^{(j)} + \sum_{k=0}^{j-1} H_{\mathbf{x}}^k D(\mathbf{x}_{t-1-k}, \mathbf{z}_{t-k}).$$

Note that $A^{(j)}$ and $\boldsymbol{\delta}^{(j)}$ are nonempty since we chose t as small as possible. Combining this with (21) gives

$$H_{\mathbf{x}}^j \mathbf{x}_{t-j} = \mathbf{x}^* - \sum_{k=0}^{j-1} H_{\mathbf{x}}^k H_0 - A^{(j)}\boldsymbol{\delta}^{(j)} + \sum_{k=j}^{t-1} H_{\mathbf{x}}^k D(\mathbf{x}_{t-1-k}, \mathbf{z}_{t-k}).$$

It then follows, since $\|\mathbf{x}_{t-j}\| \geq \|H_{\mathbf{x}}^j\|^{-1} \|H_{\mathbf{x}}^j \mathbf{x}_{t-j}\|$, that we get

(22)
$$\|\mathbf{x}_{t-j}\| \geq \|H_{\mathbf{x}}^j\|^{-1} \left(\left\| A^{(j)}\boldsymbol{\delta}^{(j)} \right\| - \|\mathbf{x}^*\| - \left\| \sum_{k=0}^{j-1} H_{\mathbf{x}}^k H_0 \right\| - \left\| \sum_{k=j}^{t-1} H_{\mathbf{x}}^k D(\mathbf{x}_{t-1-k}, \mathbf{z}_{t-k}) \right\| \right).$$

The remaining part of the proof is a recursive argument. We start at $j = t-1$, in which case (22) gives

$$\|\mathbf{x}_1\| \geq d_1 \left(\left\| A^{(t-1)}\boldsymbol{\delta}^{(t-1)} \right\| - \|D(\mathbf{x}_0, \mathbf{z}_1)\| \right) + d_2.$$

This goes to infinity linearly in d as we increase d, as the first norm increases linearly with d while

$$\lim_{d \to \infty} D(\mathbf{x}_0, \mathbf{z}_1) = 0,$$

because the deviation is exponentially fast decreasing in its first argument and increasing at only a polynomial rate in its second argument. Next, since $\|\mathbf{x}_1\|$ goes to infinity linearly in d, it follows by a similar argument

$$\|\mathbf{x}_2\| \geq d_2 \left(\left\| A^{(t-2)}\boldsymbol{\delta}^{(t-2)} \right\| - \|D(\mathbf{x}_1, \mathbf{z}_2) + H_{\mathbf{x}} D(\mathbf{x}_0, \mathbf{z}_1)\| \right) + d_3$$

goes to infinity linearly in d as we increase d. Iterate until $\mathbf{x}_{t-1}$ to conclude that each $\|\mathbf{x}_{t-j}\|$ for $0 < j \leq t$ increases linearly with d to infinity and thus we can always choose d large enough to ensure that the deviations in (21) are arbitrarily close to zero.

A.2 Proofs of Section 4

A.2.1 Proof of Lemma 4.1

We can rewrite

$$\sum_{k=0}^{\infty} \rho(H_{\mathbf{x}})^k \delta_{t-k} = (1 - \rho(H_{\mathbf{x}})) \sum_{k=0}^{\infty} \delta_{t-k} \sum_{j=k}^{\infty} \rho(H_{\mathbf{x}})^j = (1 - \rho(H_{\mathbf{x}})) \sum_{j=0}^{\infty} \rho(H_{\mathbf{x}})^j \sum_{k=0}^{j} \delta_{t-k}.$$

Next, $(\delta_t)_{t\in\mathbb{Z}}$ is a stationary ergodic sequence by Krengel's lemma, Proposition 4.3 in Krengel (1985), and $\mathbb{E}\delta_{t-k} < \infty$ by the assumption that $\mathbb{E}\|\boldsymbol{\varepsilon}_t\|^m < \infty$ and part (ii) of Theorem 3.2. Therefore a law of large numbers holds and thus

$$\begin{aligned}\lim_{\rho(H_{\mathbf{x}})\to 1}(1-\rho(H_{\mathbf{x}}))\sum_{k=0}^{\infty}\rho(H_{\mathbf{x}})^k\delta_{t-k} &= \lim_{\rho(H_{\mathbf{x}})\to 1}(1-\rho(H_{\mathbf{x}}))^2\sum_{j=0}^{\infty}\rho(H_{\mathbf{x}})^j\sum_{k=0}^{j}\delta_{t-k}\\ &= \lim_{\rho(H_{\mathbf{x}})\to 1}(1-\rho(H_{\mathbf{x}}))^2\sum_{j=0}^{\infty}\rho(H_{\mathbf{x}})^j(j+1)\mathbb{E}\delta_{t-k}\\ &= \mathbb{E}\delta_0.\end{aligned}$$

A.3 Proofs of Section 5

A.3.1 Proof of Proposition 5.1

Note that

$$\|h_{tp}^{(m)}(\mathbf{x},\mathbf{z},\sigma)-h(\mathbf{x},\mathbf{z},\sigma)\| \leq \|h_{tp}^{(m)}(\mathbf{x},\mathbf{z},\sigma)-h_p^{(m)}(\mathbf{x},\mathbf{z},\sigma)\| + \|h_p^{(m)}(\mathbf{x},\mathbf{z},\sigma)-h(\mathbf{x},\mathbf{z},\sigma)\|.$$

Now,

$$\lim_{m\to\infty}\sup_{(\mathbf{x},\mathbf{z},\sigma)\in S}\|h_{tp}^{(m)}(\mathbf{x},\mathbf{z},\sigma)-h_p^{(m)}(\mathbf{x},\mathbf{z},\sigma)\| = 0,$$

because S is compact and $\tau\to 0$ as $m\to\infty$ and

$$\lim_{m\to\infty}\sup_{(\mathbf{x},\mathbf{z},\sigma)\in S}\|h_p^{(m)}(\mathbf{x},\mathbf{z},\sigma)-h(\mathbf{x},\mathbf{z},\sigma)\| = 0,$$

by the assumptions that the true policy function is analytic over a compact set S and the Weierstrass M-test.

A.3.2 Proof of Proposition 5.2

This result follows immediately by noticing that setting $\tau = 0$ makes the transformed polynomials equal to the regular polynomials. Therefore we can always find a τ for which transformed perturbation performs equally or better than regular perturbation.

A.3.3 Proof of Proposition 5.3

Let $(\bar{\mathbf{x}}_t)_{t\geq 0}$ be the path generated by the m'th order perturbation policy function, also initialised at the origin. Additionally, let $\mathbf{v}_t = (\mathbf{x}_{t-1},\mathbf{z}_t)$ and $\bar{\mathbf{v}}_t = (\bar{\mathbf{x}}_{t-1},\mathbf{z}_t)$. Throughout this proof we let $\|\cdot\|$ be the infinity norm, or maximum norm.

It follows from the exogenous variable updating function in (2) and the fact that $\mathbf{z}_0 = \mathbf{0}_{n_z}$ that

$$\|\mathbf{z}_t\| \leq \|\boldsymbol{\Lambda}\|\|\mathbf{z}_{t-1}\| + \sigma\|\boldsymbol{\eta}\boldsymbol{\varepsilon}_t\| = \|\boldsymbol{\Lambda}\|\|\mathbf{z}_{t-1}\| + O(\sigma) = \|\boldsymbol{\Lambda}\|^t\|\mathbf{z}_0\| + O(\sigma) = O(\sigma), \qquad \forall t\in\mathbb{N}.$$

Next, we proof by induction that $\|\bar{\mathbf{x}}_t\| = O(\sigma)$ for all $t\in\mathbb{N}$. It is true for $t=1$, since $\mathbf{x}_0 = \mathbf{0}_{n_x}$ and thus

$$\begin{aligned}\|\bar{\mathbf{x}}_1\| &\leq \|H_0\| + \|H_{\mathbf{z}}\|\|\mathbf{z}_1\| + \sum_{i=2}^{m}\|H_i\|\|\bar{\mathbf{v}}_1\|^i\\ &= O(\sigma) + O(\sigma) + \sum_{i=2}^{m}\|H_i\|\|\mathbf{z}_1\|^i = O(\sigma),\end{aligned}$$

where we used that $\|\mathbf{z}_1\| = O(\sigma)$ by the previous derivation and $\|H_0\| = O(\sigma)$ by the definition of H_0. Similarly, if $\|\bar{\mathbf{x}}_{t-1}\| = O(\sigma)$, then

$$\|\bar{\mathbf{x}}_t\| \leq \|H_0\| + \|H_{\mathbf{x}}\|\|\bar{\mathbf{x}}_{t-1}\| + \|H_{\mathbf{z}}\|\|\mathbf{z}_t\| + \sum_{i=2}^{m} \|H_i\|\|\bar{\mathbf{v}}_t\|^i = O(\sigma).$$

We proceed by showing via induction that $\|\bar{\mathbf{x}}_t - \mathbf{x}_t\| = O(\sigma^{m+1})$ and $\|\mathbf{x}_t\| = O(\sigma)$ for all $t \in \mathbb{N}$. This is true for $t = 1$, since by the reverse triangle inequality and the properties of a Taylor approximation we have that

$$\begin{aligned} |\|\bar{\mathbf{x}}_1\| - \|\mathbf{x}_1\|| \leq \|\bar{\mathbf{x}}_1 - \mathbf{x}_1\| &= \|h_p(\mathbf{v}_1, \sigma) - h(\mathbf{v}_1, \sigma)\| \\ &= O\left(\|(\mathbf{v}_1, \sigma)\|^{m+1}\right) = O\left(\|(\mathbf{z}_1, \sigma)\|^{m+1}\right) = O\left(\sigma^{m+1}\right). \end{aligned}$$

If the statement hold for $t-1$, then likewise

$$\begin{aligned} |\|\bar{\mathbf{x}}_t\| - \|\mathbf{x}_t\|| \leq \|\bar{\mathbf{x}}_t - \mathbf{x}_t\| &= \|h_p(\bar{\mathbf{v}}_t, \sigma) - h(\mathbf{v}_t, \sigma)\| \\ &\leq \|h_p(\bar{\mathbf{v}}_t, \sigma) - h_p(\mathbf{v}_t, \sigma)\| + \|h_p(\mathbf{v}_t, \sigma) - h(\mathbf{v}_t, \sigma)\| \end{aligned}$$

The second term is of $O\left(\sigma^{m+1}\right)$ by the same argument as before. The first term requires a bit more work

$$\|h_p(\bar{\mathbf{v}}_t, \sigma) - h_p(\mathbf{v}_t, \sigma)\| \leq \|H_{\mathbf{x}}\|\|\bar{\mathbf{x}}_{t-1} - \mathbf{x}_{t-1}\| + \sum_{i=2}^{m} \|H_i\| \left\| \bigotimes_i \bar{\mathbf{v}}_t - \bigotimes_i \mathbf{v}_t \right\|,$$

which is $O\left(\sigma^{m+1}\right)$ since

$$\begin{aligned} \left\| \bigotimes_i \bar{\mathbf{v}}_t - \bigotimes_i \mathbf{v}_t \right\| &\leq i\|\bar{\mathbf{v}}_t - \mathbf{v}_t\| \max\{\|\bar{\mathbf{v}}_t\|, \|\mathbf{v}_t\|\}^{i-1} \\ &= i\|\bar{\mathbf{x}}_{t-1} - \mathbf{x}_{t-1}\|\|(\bar{\mathbf{v}}_t, \mathbf{v}_t)\|^{i-1} = O\left(\sigma^{m+1}\right). \end{aligned}$$

The next step is to show that $\|\bar{\mathbf{x}}_t - \hat{\mathbf{x}}_t\| = O\left(\sigma^{\min\{m+1,4\}}\right)$ and $\|\hat{\mathbf{x}}_t\| = O(\sigma)$ for all $t \in \mathbb{N}$. Since $\mathbf{x}_0 = \mathbf{0}_{n_x}$ we have $\|\bar{\mathbf{x}}_1 = \hat{\mathbf{x}}_1\|$. Let $\hat{\mathbf{v}}_t = (\hat{\mathbf{x}}_{t-1}, \mathbf{z}_t)$ and suppose the statement holds for $t-1$, then similarly as before we have

$$\begin{aligned} |\|\bar{\mathbf{x}}_t\| - \|\hat{\mathbf{x}}_t\|| \leq \|\bar{\mathbf{x}}_t - \hat{\mathbf{x}}_t\| &\leq \|H_{\mathbf{x}}\|\|\bar{\mathbf{x}}_{t-1} - \hat{\mathbf{x}}_{t-1}\| \\ &\quad + \sum_{i=2}^{m} \|H_i\| \left\| \bigotimes_i \bar{\mathbf{v}}_{t-1} - \bigotimes_i \hat{\mathbf{v}}_{t-1} \right\| + \sum_{i=2}^{m} \|H_i\|\|\hat{\mathbf{v}}_{t-1}\|^{i-1} \left\| 1 - \Phi_\tau\left(\tilde{\hat{\mathbf{x}}}_{t-1}\right) \right\| \\ &= O\left(\sigma^{\min\{m+1,4\}}\right) + O\left(\sigma^{\min\{m+1,4\}}\right) + O\left(\sigma^2\right) \left\| 1 - \Phi_\tau\left(\tilde{\hat{\mathbf{x}}}_{t-1}\right) \right\|. \end{aligned}$$

Note that

$$\left\| 1 - \Phi_\tau\left(\tilde{\hat{\mathbf{x}}}_{t-1}\right) \right\| = \left\| 1 - e^{-\tau\|\tilde{\hat{\mathbf{x}}}_{t-1}\|_e^2} \right\| = O(\|\hat{\mathbf{x}}_{t-1}\|^2) = O\left(\sigma^2\right),$$

so that the result follows. The proposition is now proved by putting everything together:

$$\|\hat{\mathbf{x}}_t - \mathbf{x}_t\| \leq \|\hat{\mathbf{x}}_t - \bar{\mathbf{x}}_t\| + \|\bar{\mathbf{x}}_t - \mathbf{x}_t\| = O\left(\sigma^{\min\{m+1,4\}}\right) + O\left(\sigma^{m+1}\right) = O\left(\sigma^{\min\{m+1,4\}}\right).$$

A.3.4 Proof of Proposition 5.4

We start by showing that condition (14) ensures that the true policy function produces nonexplosive sample paths. This follows from Theorem 9.4.1 in Meyn and Tweedie (1993), which states that we have to find a non-negative function $V : \mathcal{X} \times \mathcal{Z} \to \mathbb{R}$ such that

$$\frac{\mathbb{E}\left(V(\mathbf{x}_1, \mathbf{z}_1) \mid \mathbf{x}_0 = \mathbf{x}, \mathbf{z}_0 = \mathbf{z}\right)}{V(\mathbf{x}, \mathbf{z})} < 1 \tag{23}$$

for all $\mathbf{x}$ and $\mathbf{z}$ outside of a compact $C \subseteq \mathcal{X} \times \mathcal{Z}$. We use the function $V(\mathbf{x}, \mathbf{z}) = \|\mathbf{x}\| + \|\mathbf{z}\|$ and obtain similarly to the proof of Theorem 3.2 that there exists a constant d such that (23) is bounded by

$$\frac{\mathbb{E}(\|h(\mathbf{x}, \mathbf{z}_1, \sigma)\| \mid \mathbf{z}_0 = \mathbf{z}) + \|\mathbf{\Lambda}\|\|\mathbf{z}\|}{\|\mathbf{x}\| + \|\mathbf{z}\|} + \frac{d}{\|\mathbf{x}\| + \|\mathbf{z}\|}.$$

Increasing $\mathbf{x}$ can make the first fraction smaller than one by condition (14), while the second fraction can be made arbitrarily small. Therefore there exists an $M > 0$ such that (23) is satisfied for all $\mathbf{x}, \mathbf{z} > M$.

Next we show that condition (14) implies (15). Note that condition (14) and Assumption C3 imply that

$$\limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h(\mathbf{x}, \mathbf{z}, \sigma)\|}{\|\mathbf{x}\|} < \infty$$

for all possible values of $\mathbf{z}$ and σ outside of a set of Lebesque measure zero. However, since $h_p(\mathbf{x}, \mathbf{z}, \sigma)$ contains a nonzero higher order monomial in $\mathbf{x}$ we have

$$\liminf_{\|\mathbf{x}\| \to \infty} \frac{\|h_p(\mathbf{x}, \mathbf{z}, \sigma)\|}{\|\mathbf{x}\|} = \infty.$$

for all nonzero values of $\mathbf{z}$ and σ. Finally, since the deviations in transformed perturbation go to zero away from the steady state we have

$$\limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h_{tp}(\mathbf{x}, \mathbf{z}, \sigma)\|}{\|\mathbf{x}\|} = \|H_{\mathbf{x}}\| < \infty.$$

It immediately follows that the difference between the true and the perturbed policy functions become infinitely many times larger than the errors between the true and the transformed perturbation policy functions as $\|\mathbf{x}\|$ goes to infinity.

In the last part we show that conditions (i), (ii) and (iii) imply (14). Condition (i) follows from the reverse Fatou lemma as

$$\limsup_{\|\mathbf{x}\| \to \infty} \frac{\mathbb{E}(\|h(\mathbf{x}, \mathbf{z}_1, \sigma)\| \mid \mathbf{z}_0 = \mathbf{z})}{\|\mathbf{x}\|} < \mathbb{E}\left(\limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h(\mathbf{x}, \mathbf{z}_1, \sigma)\|}{\|\mathbf{x}\|} \middle| \mathbf{z}_0 = \mathbf{z}\right) < 1.$$

Condition (ii) immediately implies condition (i) and condition (iii) implies condition (i) since

$$\begin{aligned}
\limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h(\mathbf{x}, \mathbf{z}, \sigma)\|}{\|\mathbf{x}\|} &= \limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h(\mathbf{a}\mathbf{x}, \mathbf{z}, \sigma)\|}{\|\mathbf{a}\mathbf{x}\|} \leq \limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h(\mathbf{a}\mathbf{x}, \mathbf{z}, \sigma)\|}{\|h(\mathbf{x}, \mathbf{z}, \sigma)\|} \limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h(\mathbf{x}, \mathbf{z}, \sigma)\|}{\|\mathbf{a}\mathbf{x}\|} \\
&= \frac{1}{a} \limsup_{\|\mathbf{x}\| \to \infty} \frac{\|h(\mathbf{x}, \mathbf{z}, \sigma)\|}{\|\mathbf{x}\|} = 0.
\end{aligned}$$

References

Andreasen, M. M., Fernández-Villaverde, J., and Rubio-Ramírez, J. (2017). The pruned state-space system for non-linear dsge models: Theory and empirical applications. *Review of Economic Studies.*

Apostol, T. (1967). *Calculus.* Wiley & Sons.

Aruoba, S. B., Fernandez-Villaverde, J., and Rubio-Ramirez, J. F. (2006). Comparing solution methods for dynamic equilibrium economies. *Journal of Economic dynamics and Control*, 30(12):2477–2508.

Bergin, P. R., Shin, H.-C., and Tchakarov, I. (2007). Does exchange rate variability matter for welfare? a quantitative investigation of stabilization policies. *European Economic Review*, 51(4):1041–1058.

Bertsekas, D. P. (1987). *Dynamic Programming: Deterministic and Stochastic Models.* Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

Blanchard, O. J. and Kahn, C. M. (1980). The solution of linear difference models under rational expectations. *Econometrica: Journal of the Econometric Society*, pages 1305–1311.

Blasques, F. (2014). Transformed polynomials for nonlinear autoregressive models of the conditional mean. *Journal of Time Series Analysis*, 35(3):218–238.

Bougerol, P. (1993). Kalman filtering with random coefficients and contractions. *SIAM Journal on Control and Optimization*, 31(4):942–959.

Christiano, L. J. and Fisher, J. D. (2000). Algorithms for solving dynamic models with occasionally binding constraints. *Journal of Economic Dynamics and Control*, 24(8):1179–1232.

Cline, D. B. and Pu, H.-m. H. (1999). Geometric ergodicity of nonlinear time series. *Statistica Sinica*, pages 1103–1118.

Davydov, Y. A. (1974). Mixing conditions for markov chains. *Theory of Probability & Its Applications*, 18(2):312–328.

Deaton Angus, S. (1991). Saving and liquidity constraints. *Ecomometrica*, 59(5):221–248.

Den Haan, W. J. and De Wind, J. (2010). How well-behaved are higher-order perturbation solutions? Technical report, Netherlands Central Bank, Research Department.

Den Haan, W. J. and De Wind, J. (2012). Nonlinear and stable perturbation-based approximations. *Journal of Economic Dynamics and Control*, 36(10):1477–1497.

Fahr, S. and Smets, F. (2010). Downward wage rigidities and optimal monetary policy in a monetary union. *The Scandinavian Journal of Economics*, 112(4):812–840.

Fernández-Villaverde, J., Guerrón-Quintana, P., Kuester, K., and Rubio-Ramírez, J. (2015). Fiscal volatility shocks and economic activity. *American Economic Review*, 105(11):3352–84.

Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramirez, J. F., and Uribe, M. (2011). Risk matters: The real effects of volatility shocks. *American Economic Review*, 101(6):2530–61.

Judd, K. L. (1992). Projection methods for solving aggregate growth models. *Journal of Economic Theory*, 58(2):410–452.

Judd, K. L. (1998). *Numerical methods in economics*. MIT press.

Judd, K. L. and Guu, S.-M. (1997). Asymptotic methods for aggregate growth models. *Journal of Economic Dynamics and Control*, 21(6):1025–1042.

Kim, J. and Kim, S. (2003a). Welfare effects of tax policy in open economies: stabilization and cooperation.

Kim, J., Kim, S., Schaumburg, E., and Sims, C. A. (2008). Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. *Journal of Economic Dynamics and Control*, 32(11):3397–3414.

Kim, J. and Kim, S. H. (2003b). Spurious welfare reversals in international business cycle models. *journal of International Economics*, 60(2):471–500.

Kollmann, R. (2002). Monetary policy rules in the open economy: effects on welfare and business cycles. *Journal of Monetary Economics*, 49(5):989–1015.

Krengel, U. (1985). *Ergodic theorems*, volume 6. Walter de Gruyter.

Kydland, F. E. and Prescott, E. C. (1982). Time to build and aggregate fluctuations. *Econometrica: Journal of the Econometric Society*, pages 1345–1370.

Lancaster, P. and Farahat, H. (1972). Norms on direct sums and tensor products. *mathematics of computation*, 26(118):401–414.

Meyn, S. P. and Tweedie, R. L. (1993). *Markov chains and stochastic stability*. Springer, London.

Plesniak, W. (1990). Markov's inequality and the existence of an extension operator for c^∞ functions. *Journal of Approximation Theory*, 61(1):106–117.

Powell, M. J. D. (1981). *Approximation theory and methods*. Cambridge university press.

Rudebusch, G. D. and Swanson, E. T. (2012). The bond premium in a dsge model with long-run real and nominal risks. *American Economic Journal: Macroeconomics*, 4(1):105–43.

Saïdi, Y. and Zakoian, J.-M. (2006). Stationarity and geometric ergodicity of a class of nonlinear arch models. *The Annals of Applied Probability*, 16(4):2256–2271.

Santos, M. S. (2000). Accuracy of numerical solutions using the euler equation residuals. *Econometrica*, 68(6):1377–1402.

Schmitt-Grohé, S. and Uribe, M. (2004). Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of economic dynamics and control*, 28(4):755–775.

Stone, M. H. (1937). Applications of the Theory of Boolean Rings to General Topology. *Transactions of the American Mathematical Society*, 41(3):375–481.

Stone, M. H. (1948). The Generalized Weiestrass Approximation Theorem. *Mathematics Magazine*, 21(4):167–184.

Straumann, D. (2005). *Estimation in Conditionally Heteroscedastic Time Series Models. Lecture Notes in Statistics*, volume 181. Springer, Berlin.

Taylor, J. B. and Uhlig, H. (1990). Solving nonlinear stochastic growth models: a comparison of alternative solution methods. *Journal of Business & Economic Statistics*, 8(1):1–17.

Tesar, L. L. (1995). Evaluating the gains from international risksharing. In *Carnegie-Rochester Conference Series on Public Policy*, volume 42, pages 95–143. Elsevier.

Tjøstheim, D. (1990). Non-linear time series and markov chains. *Advances in Applied Probability*, 22(3):587–611.

Van Binsbergen, J. H., Fernández-Villaverde, J., Koijen, R. S., and Rubio-Ramírez, J. (2012). The term structure of interest rates in a dsge model with recursive preferences. *Journal of Monetary Economics*, 59(7):634–648.

Woodford, M. (2002). Inflation stabilization and welfare. *Contributions in Macroeconomics*, 2(1).