Fehrle, Daniel; Heiberger, Christopher; Huber, Johannes

Polynomial Chaos Expansion: Efficient Evaluation and Estimation of Computational Models

Computational Economics

Provided in Cooperation with:
Springer Nature

Suggested Citation: Fehrle, Daniel; Heiberger, Christopher; Huber, Johannes (2025) : Polynomial Chaos Expansion: Efficient Evaluation and Estimation of Computational Models, Computational Economics, ISSN 1572-9974, Springer US, New York, NY, Vol. 65, Iss. 2, pp. 1083-1146, https://doi.org/10.1007/s10614-024-10772-5

Computational Economics (2025) 65:1083–1146
https://doi.org/10.1007/s10614-024-10772-5

Polynomial Chaos Expansion: Efficient Evaluation and Estimation of Computational Models

Daniel Fehrle[1] · **Christopher Heiberger**[2] · **Johannes Huber**[3]

Accepted: 25 October 2024 / Published online: 23 January 2025

Abstract
We apply Polynomial chaos expansion (PCE) to surrogate time-consuming repeated model evaluations for different parameter values. PCE represents a random variable, the quantity of interest (QoI), as a series expansion of other random variables, the inputs. Repeated evaluations become inexpensive by treating uncertain parameters of a model as inputs, and an element of a model's solution, e.g., the policy function, second moments, or the posterior kernel as the QoI. We introduce the theory of PCE and apply it to the standard real business cycle model as an illustrative example. We analyze the convergence behavior of PCE for different QoIs and its efficiency when used for estimation. The results are promising both for local and global solution methods.

Keywords Polynomial chaos expansion · Parameter inference · Parameter uncertainty · Solution methods

JEL Classification C11 · C13 · C32 · C63

✉ Johannes Huber
johannes1.huber@ur.de

Daniel Fehrle
fehrle@economics.uni-kiel.de

Christopher Heiberger
christopher.heiberger@wiwi.uni-augsburg.de

[1] Department of Economics, Kiel University, Wilhelm-Seelig-Platz 1, 24118 Kiel, Germany

[2] Department of Economics, University of Augsburg, Universitätsstraße 16, 86159 Augsburg, Germany

[3] Department of Economics, University of Regensburg, Universitätsstraße 31, 93053 Regensburg, Germany

1 Introduction

At an abstract level, computational economic models are mappings from inputs to outputs of the model. The former are the model's parameters, the latter are the quantities of interest (QoIs) and depend on the research question. Typical QoIs are the economic agents' policy functions, the second moments of the model's variables, or the likelihood implied by a given set of observed data. In all cases, the model's parameters are typically unknown, and plausible values must be derived from observed data or treated as random variables from the Bayesian perspective. Either way, the uncertainty of parameters translates into uncertainty regarding the model's outcomes. Estimation methods, such as minimum distance estimators or likelihood-based methods as well as a careful study of the sensitivity of the model's outcomes for a set of different parameter values require numerous repeated solutions of the model. Depending on the complexity of the model, estimation and sensitivity analyses can become a time-consuming computational task or even excessive. We show that PCE offers an elegant way to deal with this problem and provide MATLAB® code to ease its implementation.

In a nutshell, PCE enables the representation of a random variable—the QoI—as a series expansion of other random variables—the inputs. Our approach is to use PCE as a surrogate of the distribution of the model outcome given some parameter uncertainty. Therefore, we depict different model outcomes as QoIs (e.g., the policy function or the posteriors kernel) in terms of a series expansion of the model's uncertain parameters. Given the respective formulae, the required repeated evaluations are time-efficient compared to repeated solutions of the entire model. Without limiting the applicability for other purposes, we apply solution and estimation methods for dynamic stochastic general equilibrium (DSGE) models as we are familiar with the required techniques.

More to the point, after introducing the theory of PCE and the construction of a truncated PCE, we apply the method to the benchmark real business cycle (RBC) model, since this model is suited as an illustrative example due to its well-known and simplistic nature. We analyze the convergence behavior of the PCE of various model outcomes including the model's linear solution, a projection solution, the variables' second moments, and the impulse response function. Further, we conduct Monte Carlo experiments. We estimate the parameters from a linearized and non-linearized model using various estimation techniques, namely generalized method of moments (GMM), simulated method of moments (SMM), maximum-likelihood estimation (MLE), and Bayesian estimation (BE).

We document linear convergence behavior. Considering three unknown parameters, we find remarkably well approximations with only a few model evaluations. Suppose the model outcome, e.g., the linearized policy function, has to be evaluated for a sample of 100,000 parameter values. In that case, the PCE with truncation degree 7 provides an approximation with L^2 error of 10^{-3} while the computational time is lower by the factor 30. We extend the analysis to a higher-dimensional problem where all six model parameters are assumed unknown. As our construction of the PCE applies a tensor basis quadrature rule, the

construction suffers from the course of dimensionality. Thus, we compare full-tensor-grid quadrature rules with two remedies: sparse-grid quadrature rules and least squares. A comparison of time versus accuracy shows that these long-established sparse methods milden the course of dimensionality. Beyond our implementations, we highlight that the repeated model evaluations required for PCE are parallelizable because the parameter values are predetermined, distinct from the recursive nature of most applications. Additionally, we discuss the expanding literature on sparse PCE.

Our analysis continues with Monte Carlo experiments as in Ruge-Murcia (2007), where we gauge the quality of the model's PCE when applied to estimation. Here, we use the linearized solution of the real business cycle (RBC) model for the data-generating process and the econometric model since this procedure allows us to calculate the analytic second moments and likelihood functions. Consequently, the differences in the estimates towards the benchmark procedure of repeated solutions are solely based on the PCE approximation. The PCE based method is remarkably efficient and accurate. Estimates deviate only negligibly from the benchmark procedure and most notable, the computation time can be reduced by 99 percent for BE and by 50 percent for GMM, SMM, and MLE.

In the last step of our analysis, we stress PCE out more by gauging the quality of the non-linear model's PCE for likelihood-based estimations. For this purpose, we replicate the findings of Fernández-Villaverde and Rubio-Ramírez (2005), who have shown that non-linearity is already relevant for the estimates of our benchmark RBC model. We show that the use of PCE for the estimation of the non-linear model enhances the accuracy of the estimates considerably in comparison to repeated, linearly-solved model estimation and reduces time up to 97 percent compared to repeated globally-solved model estimation. Also worth noting, with PCE as a surrogate for the likelihood, the likelihood from a particle filter becomes continuously differentiable—allowing a gradient-based optimization.

In its general form, the underlying theory of the method rests on the theory introduced by Wiener (1938) and the Cameron and Martin (1947) theorem for a family of stochastically independent and normally distributed random variables and Hermite polynomials. The property does not only hold for Hermite polynomials and probability measures of normally distributed random variables but also extends to other commonly used distributions and the corresponding orthogonal polynomials from the Askey scheme. This extension, initially proposed by Xiu and Karniadakis (2002), is also known as generalized polynomial chaos expansion. Ghanem and Spanos (1991) provide the first applications of the theory to the problem of uncertain model parametrization.

There are two pioneer applications in economics. Pröhl (2017) uses PCE to discretize the state-space of the benchmark heterogeneous agent model and Harenberg et al. (2019) use the polynomial coefficients for global sensitivity analysis of the RBC model. Gersbach et al. (2021) follow Harenberg et al. (2019) and use PCE to identify decisive parameters. In finance PCE is e.g., applied by Albeverio et al. (2019); Dias and Peters (2021); Marconi (2016). The application of PCE in Bayesian inference was first analyzed by Marzouk et al. (2007) in engineering but to the best of our knowledge, the method has not yet been studied to estimate computational

economic models. Scheidegger and Bilionis (2019) incorporate parameter uncertainty in their solution of computational economic models by rewriting the parameters as states. This way, the model is indirectly solved in the parameter domain.

The remainder of the paper is structured as follows. First, in section 2 we review the basic theory for the existence of polynomial chaos expansions, present common practical methods to compute the PCE coefficients, and discuss the application for a pointwise approximation of the mapping from the parameters to the model outcome, i.e., the construction of a surrogate. In section 3, we apply our approach to the benchmark RBC model and discuss our results and potential drawbacks. Section 4 concludes. More detailed derivations, applications, etc. can be found in the appendix.

2 Generalized Polynomial Chaos Expansions

We begin by reviewing the basic idea and theory behind the concept of PCE. While PCE proved useful for various applications, we focus on their implementation to efficiently evaluate computationally expensive model outcomes when one or more of the model's inputs, i.e., model parameters, are uncertain. Further, we give an analytically tractable example to outline the concept of PCE in Appendix 1.

Notation and Preliminaries We consider a computational economic model where $\vartheta_i \in \Theta_i, \Theta_i \subset \mathbb{R}, i = 1, \dots, k$, denotes an arbitrary selection of $k \in \mathbb{N}$ parameters of the model. Moreover, we are interested in some model outcome(s) denoted by a vector $y \in \mathbb{R}^m, m \in \mathbb{N}$. The relation between the input parameters ϑ_i and the model outcome(s) y is determined deterministically, i.e., repeated computation of y with the same inputs ϑ_i to the model produces the same result.[1] This mapping between the ϑ_i and y is described by

$$y = h(\vartheta_1, \dots, \vartheta_k)$$

where

$$h \colon \Theta \to \mathbb{R}^m, \Theta = \bigtimes_{i=1}^{k} \Theta_i \subset \mathbb{R}^k.$$

Without loss of generality, we consider the case $m = 1$ in the following and note that for $m \geq 2$ all derivations can be applied separately to each component y_i of y, $i = 1, \dots, m$, in the same way.

Now further consider the case where the values ϑ_i of the model parameters are subject to some uncertainty to the researcher. In order to account for this uncertainty, we switch from the deterministic representation of the parameters to the perspective of

[1] E.g., if y denotes some second moments of the model, these are derived either from available analytic formulae from the (approximated) model solution or are computed from simulations with the same sample of shocks.

describing them by appropriately distributed random variables. Therefore, let $(\Omega, \mathcal{A}, P)$ denote a sufficiently rich probability space so that any uncertain model input parameter can be described by some real-valued random variable $\theta_i : \Omega \to \mathbb{R}, i = 1, \ldots, k$, where the real line is equipped with the Borel sigma-algebra $\mathcal{B}(\mathbb{R})$. Moreover, let $\xi_1, \ldots, \xi_k$ denote a family of stochastically independent random variables chosen by the researcher as a basis of the desired polynomial expansions, the so-called germs. In applications, as will be described later, the germs are most commonly either set equal to the uncertain model parameters θ_i or to some natural and convenient transformation of them. We assume:

1. The germs $\xi_1, \ldots, \xi_k$ cover the same stochastic information as the uncertain model parameters, i.e.,

$$\sigma(\xi_1, \ldots, \xi_k) = \sigma(\theta_1, \ldots, \theta_k),$$

   where $\sigma(\cdot)$ denotes the sigma-algebra generated by the random variables.
2. All moments of each ξ_i exist, i.e., $\mathbb{E}[|\xi_i|^n] < \infty$ for all $i = 1, \ldots, k$ and $n \in \mathbb{N}_0$.

Moreover, we write $\boldsymbol{\theta} := (\theta_1, \ldots, \theta_k) : \Omega \to \mathbb{R}^k$ and $\boldsymbol{\xi} := (\xi_1, \ldots, \xi_k) : \Omega \to \mathbb{R}^k$ for the k-dimensional random vector of the uncertain model parameters and for the random vector of the germs, respectively, where $\mathbb{R}^k$ is also equipped with its Borel sigma-algebra $\mathcal{B}(\mathbb{R}^k)$. For each $i = 1, \ldots, k$, let $P_{\xi_i} := P \circ \xi_i^{-1}$ denote the probability measure of ξ_i on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ and analogously let $P_{\boldsymbol{\xi}} := P \circ \boldsymbol{\xi}^{-1} = \bigotimes_{i=1}^{k} P_{\xi_i}$ denote the product probability measure of $\boldsymbol{\xi}$ on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$. The Hilbert space (of equivalence classes) of square-integrable real-valued functions on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{\xi_i})$ is denoted by

$$\begin{aligned} L_i^2 &:= L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{d}P_{\xi_i}) \\ &:= \left\{ f : \mathbb{R} \to \mathbb{R} \;\middle|\; f \text{ is measurable and } \int_{\mathbb{R}} f^2 \mathrm{d}P_{\xi_i} < \infty \right\}, \end{aligned}$$

where the inner product is defined by

$$\langle f, g \rangle_{L_i^2} := \int_{\mathbb{R}} fg \mathrm{d}P_{\xi_i} = \mathbb{E}[f(\xi_i)g(\xi_i)] \quad \text{for } f, g \in L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_{\xi_i}).$$

We use the notation $\| \cdot \|_{L_i^2}$ for the induced norm on L_i^2. We introduce the analogous notation, i.e., $L^2 := L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), \mathrm{d}P_{\boldsymbol{\xi}})$, for the space of square integrable real valued functions on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_{\boldsymbol{\xi}})$ and write $\langle \cdot, \cdot \rangle_{L^2}$ and $\| \cdot \|_{L^2}$ for the inner product and for the induced norm on L^2. If the distributions of the random variables ξ_i possess probability density functions $w_i : \mathbb{R} \to \mathbb{R}_+$, the inner products become

$$\langle f, g \rangle_{L_i^2} = \int_{\mathbb{R}} f(s)g(s)w_i(s)\mathrm{d}s,$$

and

$$\langle f, g \rangle_{L^2} = \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(s_1, \dots, s_k) g(s_1, \dots, s_k) w_1(s_1) \cdot \dots \cdot w_k(s_k) \mathrm{d}s_1 \dots \mathrm{d}s_k,$$

so that $L_i^2 = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), w_i(s)\mathrm{d}s)$ and $L^2 = L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), w(s)\mathrm{d}s)$ where w is the joint probability function $w(s) := \prod_{i=1}^k w_i(s_i)$. Note that Assumption 2 is equivalent to the fact that for each $i = 1, \dots, k$ all univariate polynomials are included in L_i^2 or, again equivalently, that all k-variate polynomials are included in L^2.

Since, by Assumption 1, each θ_i is $\sigma(\boldsymbol{\xi})$-measurable, there exist measurable $\psi_i : \mathbb{R}^k \to \mathbb{R}$ which satisfy

$$\theta_i = \psi_i \circ \boldsymbol{\xi}.$$

We write $\boldsymbol{\psi} := (\psi_1, \dots, \psi_k) : \mathbb{R}^k \to \mathbb{R}^k$ so that $\boldsymbol{\theta} = \boldsymbol{\psi} \circ \boldsymbol{\xi}$. Moreover, note that $\sigma(\boldsymbol{\xi}) = \sigma(\boldsymbol{\theta})$ also implies the existence of a measurable, inverse mapping $\boldsymbol{\psi}^{-1}$ with $\boldsymbol{\psi} \circ \boldsymbol{\psi}^{-1} = \boldsymbol{\psi}^{-1} \circ \boldsymbol{\psi} = \mathrm{id}$. A further assumption we make is that

3. the second moment of each model input parameter exists, i.e., $\mathbb{E}[\theta_i^2] < \infty$ for $i = 1, \dots, k$. Equivalently, each ψ_i is square integrable on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_{\boldsymbol{\xi}})$, i.e., $\psi_i \in L^2$ for all $i = 1, \dots, k$.[2]

Moreover, as the model input parameters θ_i are now treated as random, the model outcome of interest is random. We therefore adapt its notation to $Y : \Omega \to \mathbb{R}$. Yet, given any elementary event $\omega \in \Omega$ and corresponding realization $\theta_i(\omega)$, the mapping between the model parameters and the model outcome is still determined deterministically by $Y(\omega) = h(\theta_1(\omega), \dots, \theta_k(\omega))$, i.e.,

$$Y = h \circ \boldsymbol{\theta} = h \circ \boldsymbol{\psi} \circ \boldsymbol{\xi}, \text{ for some } h : \mathbb{R}^k \to \mathbb{R}.$$

The final assumption is that Y is a well-defined random variable with finite second moments, i.e.,

4. h is measurable and $h \circ \boldsymbol{\psi}$ is square integrable on $(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_{\boldsymbol{\xi}})$, i.e., $h \circ \boldsymbol{\psi} \in L^2$.

2.1 Single Uncertain Parameter and Germ (k=1)

We begin our description with the simplest case with only one single uncertain input parameter θ and one single germ ξ, i.e., $k = 1$. In general, any arbitrary choice of the germ that satisfies Assumption 2 implies that all polynomials are included in L^2, and therefore allows the construction of an orthogonal system of polynomials $\{q_n\}_{n \in \mathbb{N}_0} \subset L^2$, i.e., a family of polynomials where q_n is of (exact) degree n and

[2] Note that the third assumption is already implied by the second if the germs are set equal to (some polynomial transformation of) the model input parameters.

$$\langle q_n, q_m \rangle_{L^2} = \|q_n\|_{L^2}^2 \delta_{m,n} \text{ for all } m, n \in \mathbb{N}_0,$$

where $\delta_{m,n}$ denotes the Knonecker delta. This can generally be achieved by applying, e.g., the Gram-Schmidt process to the sequence of monomials.

In practice, the distribution of the uncertain input parameter is given and one is free to set the germ. It is then convenient to define the germ in such a way that i) an easy representation $\theta = \psi(\xi)$ of the parameter in terms of the germ arises and ii) the family of orthogonal polynomials in L^2 corresponds to some well-known class of polynomials. Table 1 summarizes the natural choice of the germ and the corresponding family of orthogonal polynomials when the input parameter is normal, uniform, Beta, or (inverse) Gamma distributed. More details for these classes are given in Appendix 2. Additionally, Xiu and Karniadakis (2002) provide a similar overview for discrete distributions.

In all of the cases presented in Table 1 the respective families of orthogonal polynomials $\{q_n\}_{n\in\mathbb{N}_0}$ form a complete orthogonal system, i.e., lie densely in $L^2 = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\xi) = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), w(s)\mathrm{d}s)$ where w is the corresponding probability density of ξ.[3] More generally, it follows from Riesz (1924) that $\{q_n\}_{n\in\mathbb{N}_0}$ is a complete orthogonal system in L^2 if and only if there exists no other measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which generates the same moments as P_ξ, i.e., if and only if there is no other measure μ such that

$$\int_{\mathbb{R}} s^n \mathrm{d}\mu = \int s^n \mathrm{d}P_\xi = \mathbb{E}[\xi^n] \text{ for all } n \in \mathbb{N}_0.$$

If completeness of $\{q_n\}_{n\in\mathbb{N}_0}$ in L^2 can be established, then Assumptions 3 and 4 guarantee the existence of Fourier series expansions of ψ and $h\circ\psi$ in the orthogonal polynomials, i.e., there are coefficients $\{\hat{\vartheta}_n\}_{n\in\mathbb{N}_0}$ and $\{\hat{y}_n\}_{n\in\mathbb{N}_0}$, $\hat{\vartheta}_n, \hat{y}_n \in \mathbb{R}$, so that

$$\psi = \sum_{n=0}^{\infty} \hat{\vartheta}_n q_n \text{ in } L^2 = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\xi),$$

$$h\circ\psi = \sum_{n=0}^{\infty} \hat{y}_n q_n \text{ in } L^2 = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\xi).$$

Note that identity and convergence is understood in L^2 which also implies pointwise convergence i.e., for a subsequence but not pointwise convergence.[4] Moreover, since $P_\theta = P_\xi \circ \psi^{-1}$, also $h = \sum_{n=0}^{\infty} \hat{y}_n (q_n \circ \psi^{-1})$ in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\theta)$.

Hence, the uncertain model input parameter $\theta = \psi\circ\xi$ as well as our model outcome $Y = h\circ\psi\circ\xi$ can both be expanded exactly by a polynomial series in the germ, i.e., by

[3] See Szegő (1939) for proofs of completeness.

[4] For conditions for pointwise convergence see e.g., Jackson (1941).

Table 1 Overview: common distributions and corresponding germs and orthogonal polynomials on L^2

Distribution of θ		Germ		Orthogonal polynomials
Family	Parametric	ξ	ψ	q_n
Normal	$\theta \sim \mathrm{N}(\mu, \sigma^2)$	$\xi := \frac{\theta-\mu}{\sqrt{2}\sigma}$	$\psi(s) = \mu + \sqrt{2}\sigma s$	(physicists) Hermite H_n
Uniform	$\theta \sim \mathrm{U}(0,1)$	$\xi := 2\theta - 1$	$\psi(s) = \frac{s+1}{2}$	Legendre L_n
Beta	$\theta \sim \mathrm{Beta}(\alpha, \beta)$	$\xi := 2\theta - 1$	$\psi(s) = \frac{s+1}{2}$	Jacobi $J_n^{(\beta-1,\alpha-1)}$
Gamma	$\theta \sim \mathrm{Gamma}(\alpha, \beta)$[1]	$\xi := \beta\theta$	$\psi(s) = \frac{s}{\beta}$	General Laguerre $La_n^{(\alpha-1)}$
Inverse Gamma	$\theta \sim \mathrm{Inv\text{-}Gamma}(\alpha, \beta)$[1]	$\xi := \frac{\beta}{\theta}$	$\psi(s) = \frac{\beta}{s}$	General Laguerre $La_n^{(\alpha-1)}$

[a] We use the scale-rate notation

$$\theta = \psi(\xi) = \sum_{n=0}^{\infty} \hat{\vartheta}_n q_n(\xi) \ \text{ in } L^2(\Omega, \mathcal{A}, P), \tag{1a}$$

$$Y = h(\theta) = h(\psi(\xi)) = \sum_{n=0}^{\infty} \hat{y}_n q_n(\xi) \ \text{ in } L^2(\Omega, \mathcal{A}, P). \tag{1b}$$

These series expansions are called the polynomial chaos expansions (PCE) of θ and Y with respect to the germ ξ. Moreover, orthogonality of $\{q_n\}_{n\in\mathbb{N}_0}$ implies that the Fourier coefficients are determined by

$$\hat{\vartheta}_n = \|q_n\|_{L^2}^{-2} \langle \psi, q_n \rangle_{L^2} = \|q_n\|_{L^2}^{-2} \int_{\mathbb{R}} \psi q_n \mathrm{d}P_\xi, \tag{2a}$$

$$\hat{y}_n = \|q_n\|_{L^2}^{-2} \langle h\circ\psi, q_n \rangle_{L^2} = \|q_n\|_{L^2}^{-2} \int_{\mathbb{R}} (h\circ\psi) q_n \mathrm{d}P_\xi. \tag{2b}$$

Now in practice, equations (1a, b) justify approximations of the uncertain model input parameter θ as well as of the model outcome Y by their truncated PCE, i.e., by

$$S_N(\theta) = S_N(\psi\circ\xi) := \sum_{n=0}^{N} \hat{\vartheta}_n q_n(\xi),$$

$$S_N(Y) = S_N(h\circ\psi\circ\xi) := \sum_{n=0}^{N} \hat{y}_n q_n(\xi).$$

The approximations then converge to the true random variables, $S_N(\theta) \to \theta$ and $S_N(Y) \to Y$ in L^2 as $N \to \infty$. Yet, equations (2a, b) from which the coefficients are defined can in general not be evaluated analytically. This involves a second

approximation for the coefficients $\hat{\vartheta}_n$ and $\hat{y}_n$. The literature on PCE provides a variety of approaches for this task, from which we want to review the most popular ones.

2.1.1 Polynomial Chaos Expansion of the Model Parameters

Since the germ can be chosen in any desired way that satisfies Assumptions 1 and 2, the following two opposing approaches can be pursued for its specification.

In the first approach, one directly fixes the transformation ψ between the uncertain model parameter and the germ. The germ's distribution then follows from the given distribution of the uncertain input parameter and the chosen definition of ψ. In principle any choice of ψ which satisfies Assumption 2 is possible. One could then construct the family of orthogonal polynomials from the germ's distribution and the expansion coefficients could be derived by numerical integration of (2a) up to any desired order. However, it is typically far more convenient to choose ψ as a simple linear transformation between the uncertain model parameter and the germ which results in a family of well-known orthogonal polynomials in L^2, see e.g., Table 1. In this case the expansion (1a) collapses to

$$\theta = \psi(\xi) = \hat{\vartheta}_0 + \hat{\vartheta}_1 q_1(\xi)$$

and the expansion coefficients $\hat{\vartheta}_0$ and $\hat{\vartheta}_1$ are already known exactly.

Conversely, the second approach fixes the distribution of the germ and constructs ψ in such a way that it is compatible with the given distribution of the uncertain parameter. This can be achieved as follows. Let F_ξ denote the desired (cumulative) distribution function of ξ and F_θ the given distribution function of θ. Then setting the germ to[5]

$$\xi := F_\xi^{-1} \circ F_\theta \circ \theta$$

yields the desired distribution for ξ. Conversely,

$$\psi = F_\theta^{-1} \circ F_\xi$$

and the expansion coefficients can again be computed from (2a) by numerical integration.

2.1.2 Polynomial Chaos Expansion of the Model Outcome

While the expansion of the model's parameter can be controlled directly by an appropriate choice of the germ, the expansion of the QoI requires some model evaluations. We present here two approaches that treat the economic computational model behind as a black box. An intrusive approach, stochastic Galerkin, can be found in Appendix 3.

[5] We denote by F^{-1} the quantile function.

Spectral Projection

The first approach derives the polynomial chaos coefficients $\hat{y}_n$ by applying numerical integration methods to (2b). For example, if ξ possesses a probability density function w, then (2b) becomes

$$\hat{y}_n = \|q_n\|_{L^2}^{-2} \int_{\mathbb{R}} h(\psi(s)) q_n(s) w(s) \mathrm{d}s.$$

Hence, a Gauss-quadrature with M nodes that corresponds to the weight function w and to the orthogonal polynomials $\{q_n\}_{n\in\mathbb{N}_0}$ yields

$$\hat{y}_n \approx \|q_n\|_{L^2}^{-2} \sum_{i=1}^{M} h(\psi(s_i)) q_n(s_i) \omega_i \approx \|q_n\|_{L^2}^{-2} \sum_{i=1}^{M} h\left(\sum_{m=1}^{N} \hat{\vartheta}_m q_m(s_i) \right) q_n(s_i) \omega_i, \quad (3)$$

where s_i and ω_i denote the quadrature's nodes and weights, respectively. The Gauss-quadrature rule with M nodes will require to evaluate the model outcome $h(\psi(s_i)) \approx h\left(\sum_{m=1}^{N} \hat{\vartheta}_m q_m(s_i) \right)$ at each of the M nodes. Since the quadrature rule with M is exact for polynomials up to degree $2M-1$, the number of nodes should be chosen appropriately. More specifically, if $h \circ \psi$ is assumed to be well approximated by its truncated partial sum $S_N(h\circ\psi)$ of degree N, the integrand, i.e., $h(\psi(s))q_n(s)$, is well approximated by polynomials of degree not larger than $2N$ for each $n = 1, \dots, N$. Hence, it should then hold that $M \geq N+1$.

Least Squares

The second approach treats the ignored higher terms $\epsilon := \sum_{n=N+1}^{\infty} \hat{y}_n q_n(\xi)$ of the truncated PCE as the residual in a linear regression

$$Y = h(\psi(\xi)) = \sum_{n=0}^{N} \hat{y}_n q_n(\xi) + \epsilon.$$

One can then either draw $M \in \mathbb{N}$ i.i.d. sample points $s_j, j = 1, \dots, M$, from the distribution P_ξ or select them according to regression design principles. After computing the corresponding model outcomes $Y_j = h(\psi(s_j)) \approx h\left(\sum_{m=1}^{N} \hat{\vartheta}_m q_m(s_j) \right)$ the expansion coefficients are determined from

$$(\hat{y}_0, \dots, \hat{y}_n) = \underset{\hat{y}_0,\dots,\hat{y}_N}{\operatorname{argmin}} \sum_{j=1}^{M} \left(Y_j - \sum_{n=0}^{N} \hat{y}_n q_n(s_j) \right)^2.$$

The number of the sample (design) points is recommended to be set twice or three times as large as the number of unknown PCE coefficients in the literature, i.e., to $M = 2(N+1)$ or $M = 3(N+1)$.

2.2 Multiple Uncertain Input Parameters ($k \geq 2$)

We now turn to the case where more than one input parameter is uncertain and where more than one germ is used in the polynomial expansions. In brief, the stochastic independence of the germs allows us to apply the procedure from the one-dimensional case to each of the finitely many dimensions.

Since Assumption 2 guarantees that all polynomials are included in each L_i^2, one can again apply the Gram-Schmidt process to the sequence of monomials and construct for each $i = 1, \ldots, k$ an orthogonal system of polynomials $\{q_{in}\}_{n\in\mathbb{N}_0} \subset L_i^2$ where q_{in} is a polynomial of (exact) degree n and

$$\langle q_{in}, q_{im}\rangle_{L_i^2} = \|q_{in}\|_{L_i^2}^2 \delta_{m,n} \quad \text{for all } m, n \in \mathbb{N}_0.$$

For any multi-index $\alpha = (\alpha_1, \ldots, \alpha_k) \in \mathbb{N}_0^k$ we define the multivariate polynomial

$$q_\alpha(\boldsymbol{\xi}) := \prod_{i=1}^{k} q_{i\alpha_i}(\xi_i).$$

Since stochastic independence of the ξ_i implies that $P_{\boldsymbol{\xi}} = \bigotimes_{i=1}^k P_{\xi_i}$, the family of multivariate polynomials $\{q_\alpha\}_{\alpha\in\mathbb{N}_0^k}$ then forms an orthogonal system in L^2. Moreover, if for each $i = 1, \ldots, k$ the orthogonal system $\{q_{in}\}_{n\in\mathbb{N}_0}$ is complete in L_i^2, then $\{q_\alpha\}_{\alpha\in\mathbb{N}_0^k}$ is also complete in L^2. In particular, this is satisfied if each θ_i is distributed according to one of the distributions specified in Table 1 and if the germs ξ_i are set accordingly. Then, since $\psi_i \in L^2$ (Assumption 3) and $h\circ\psi \in L^2$ (Assumption 4), there exist coefficients $\{\hat{\vartheta}_{i\alpha}\}_{\alpha\in\mathbb{N}_0^k} \subset \mathbb{R}$, $i = 1, \ldots, k$, and $\{\hat{y}_\alpha\}_{\alpha\in\mathbb{N}_0^k} \subset \mathbb{R}$ such that

$$\psi_i = \sum_{\alpha\in\mathbb{N}_0^k} \hat{\vartheta}_{i\alpha} q_\alpha \text{ in } L^2 = L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_{\boldsymbol{\xi}}), \tag{4a}$$

$$h\circ\psi = \sum_{\alpha\in\mathbb{N}_0^k} \hat{y}_\alpha q_\alpha \text{ in } L^2 = L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_{\boldsymbol{\xi}}). \tag{4b}$$

The second expansion can again be written equivalently as

$$h = \sum_{\alpha\in\mathbb{N}_0^k} \hat{y}_\alpha (q_\alpha \circ \psi^{-1}) \text{ in } L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_{\boldsymbol{\theta}}).$$

Therefore, the parameters θ_i and the model outcome Y are again representable in L^2 by a PCE in the germs $\boldsymbol{\xi}$ through

$$\theta_i = \psi_i \circ \boldsymbol{\xi} = \sum_{\alpha\in\mathbb{N}_0^k} \hat{\vartheta}_{i\alpha} q_\alpha(\boldsymbol{\xi}) \text{ in } L^2(\Omega, \mathcal{A}, P), \tag{5a}$$

$$Y = h \circ \boldsymbol{\theta} = h \circ \psi \circ \boldsymbol{\xi} = \sum_{\alpha \in \mathbb{N}_0^k} \hat{y}_\alpha q_\alpha(\boldsymbol{\xi}) \text{ in } L^2(\Omega, \mathcal{A}, P). \tag{5b}$$

Moreover, the expansion coefficients are determined by

$$\hat{\vartheta}_{i\alpha} = \|q_\alpha\|_{L^2}^{-2} \langle \psi_i, q_\alpha \rangle_{L^2} = \|q_\alpha\|_{L^2}^{-2} \int_{\mathbb{R}^k} \psi_i q_\alpha \mathrm{d}P_{\boldsymbol{\xi}}, \tag{6a}$$

$$\hat{y}_\alpha = \|q_\alpha\|_{L^2}^{-2} \langle h \circ \psi, q_\alpha \rangle_{L^2} = \|q_\alpha\|_{L^2}^{-2} \int_{\mathbb{R}^k} (h \circ \psi) q_\alpha \mathrm{d}P_{\boldsymbol{\xi}}, \tag{6b}$$

where $P_{\boldsymbol{\xi}} = \bigotimes_{i=1}^k P_{\xi_i}$ implies that $\|q_\alpha\|_{L^2} = \prod_{i=1}^k \|q_{i\alpha_i}\|_{L_i^2}$.

Equations (6a, b) guarantee that if the parameters θ_i and the model outcome Y are approximated by their truncated PCE, the approximations converge to the true random variables in L^2 as the degree of the partial sums is increased. The truncation is typically introduced either by limiting the total degree of the multivariate polynomials

$$S_N^{\mathrm{tot}}(\theta_i) = S_N^{\mathrm{tot}}(\psi_i \circ \boldsymbol{\xi}) := \sum_{\alpha \in \mathbb{N}_0^k, |\alpha| \leq N} \hat{\vartheta}_{i\alpha} q_\alpha(\boldsymbol{\xi}), \tag{7a}$$

$$S_N^{\mathrm{tot}}(Y) = S_N^{\mathrm{tot}}(h \circ \psi \circ \boldsymbol{\xi}) := \sum_{\alpha \in \mathbb{N}_0^k, |\alpha| \leq N} \hat{y}_\alpha q_\alpha(\boldsymbol{\xi}), \tag{7b}$$

where $|\alpha| := \sum_{i=1}^k \alpha_i$, or by limiting the maximal degree in each component

$$S_N^{\max}(\theta_i) = S_N^{\max}(\psi_i \circ \boldsymbol{\xi}) := \sum_{\alpha \in \mathbb{N}_0^k, \|\alpha\|_\infty \leq N} \hat{\vartheta}_{i\alpha} q_\alpha(\boldsymbol{\xi}), \tag{8a}$$

$$S_N^{\max}(Y) = S_N^{\max}(h \circ \psi \circ \boldsymbol{\xi}) := \sum_{\alpha \in \mathbb{N}_0^k, \|\alpha\|_\infty \leq N} \hat{y}_\alpha q_\alpha(\boldsymbol{\xi}), \tag{8b}$$

where $\|\alpha\|_\infty := \max_{i=1,\ldots,k} \alpha_i$.

In order to compute the expansion coefficients from their defining equations (6a, b), it is straightforward to adapt the methods from section 2.1.2 to the multi-dimensional case. However, this typically introduces the curse of dimensionality.

First, this issue becomes particularly problematic if Gauss-quadrature rules compute the integrals. If the mapping $h \circ \psi$ can be well approximated by its truncated series expansion S_N, then the integrands $(h \circ \psi) q_\alpha$ in (6b) can be well approximated by multivariate polynomials which rise to degree $2N$ in each component, indifferent from the fact whether $|\alpha| \leq N$ or $\|\alpha\|_\infty \leq N$ is assumed. Since one-dimensional Gauss-quadrature rules with M nodes provide exact integration rules for polynomials up to degree $2M - 1$, it is required to compute (6b) by quadrature rules with $M = N + 1$ nodes in each of the k dimensions. Hence, the model

outcome must be evaluated for a total of $(N+1)^k$ parameter combinations and the procedure becomes quickly inefficient as k rises.

However, sparse grid methods, e.g., Smolyak-Gauss or monomial quadrature rules can help to reduce the computational effort for similar integration quality. The Smolyak-Gauss quadrature is illustrated in Appendix 4 and analyzed in the numerical example in section 3. The numerical examples for Monomial rules are presented in Appendix 5. We put them in the Appendix due to the absence of suitable rules for practical purposes. Well-performing rules are missing for truncation levels $N > 4$, dimensions $k > 6$, or mixed or non-uniform-non-normal distributions (see Stroud, 1971; Adurthi et al, 2018; Bhusal & Subbarao, 2020) However, we find remarkable results for the few suitable cases, which motivates further research to find high-degree, high-dimensional monomial rules for mixed distributions.

Second, the burden of higher-dimensional parameter vectors appears similarly if least squares determine the PCE coefficients. However, while the number of coefficients that must be computed equals $(N+1)^k$ in $S_N^{\max}$, the number of coefficients grows less extremely in S_N^{tot} where it is given by $\binom{N+k}{k}$. Following the recommendation that the number of sample points should be twice or three times as large as the number of unknown coefficients, the model must be evaluated for $2\binom{N+k}{k}$ or $3\binom{N+k}{k}$ parameter combinations in the latter case.

Sparse-grid methods and least squares give fundamentals for a rising number of more efficient alternatives. Kaintura et al. (2018) and Harenberg et al. (2019) give short discussions on recent developments of sparse-grid methods in the PCE context, whereas Lüthen et al. (2021) provide a full survey of sparse PCE. An example is an adaptive sparse grid, e.g., by eliminating points with Bayesian shrinkage priors or non-significant bases of a regression (Bürkner et al., 2023; Cheng and Lu, 2018).

2.3 Using the Expansion as Pointwise Approximation for the Model Outcome

After its construction, the PCE of the model outcome can be used for several use cases. We present common applications in Appendix 6 and focus here on our application, a pointwise approximation for the mapping between model inputs and any model outcome (e.g., the model solution - in the form of its policy function -, the second moments or the likelihood function).

More to the point, a truncated version of the Fourier series expansion (4b) can be used as a pointwise approximation for the mapping h between model parameters and a QoI

$$h(\vartheta) \approx S_N(h \circ \psi)(\psi^{-1}(\vartheta)) = \sum_{\alpha \in \mathbb{N}_0^k, |\alpha| \leq N} \hat{y}_\alpha q_\alpha(\psi^{-1}(\vartheta)). \tag{9}$$

Note, however, that convergence of the series in L^2 as $N \to \infty$ does not imply pointwise convergence on the support of P_ξ but only pointwise convergence i.e., for a subsequence.

The partial sum $S_N(h \circ \psi)$ is the orthogonal projection of $h \circ \psi$ onto the subspace of $L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_\xi)$ spanned by multivariate polynomials of total degree less or equal to N. If the transformation ψ between germs and parameters is chosen linear, $S_N(h \circ \psi) \circ \psi^{-1}$ is also the orthogonal projection of h onto this subspace in $L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_\theta)$.[6] In the sense of the induced metric, it is, therefore, the best approximation of h by multivariate polynomials of total degree up to N, i.e., it minimizes the mean-squared error over the support of P_θ.

Special Case: Surrogate of Model Solution Consider a discretely-timed model where in any period $t \in \mathbb{N}$ the vector $x_t \in S \subset \mathbb{R}^{n_x}$ denotes the predetermined variables from the state space S and $y_t \in \mathbb{R}^{n_y}$ is a vector of the non-predetermined variables of the model. Suppose that $\boldsymbol{\theta}$ is a random vector of unknown parameters of the model, and for any possible realization $\vartheta \in \Theta$ the model solution is computed in terms of a policy function $g(.;\vartheta) : S \to \mathbb{R}^{n_x + n_y}$ so that

$$\begin{pmatrix} x_{t+1} \\ y_t \end{pmatrix} = g(x_t; \vartheta).$$

If, for any arbitrary $x \in S$ and a suitable transformation ψ between parameters and germs, the mapping $\vartheta \mapsto g(x;\vartheta)$ satisfies the sufficient condition in assumption 4, then there exists a series expansion by orthogonal polynomials $\{q_\alpha\}$ of the form

$$g(x, \vartheta) = \sum_{\alpha \in \mathbb{N}_0^k} \hat{g}_\alpha(x) q_\alpha(\psi^{-1}(\vartheta)) \text{ in } L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_\theta),$$

$$\hat{g}_\alpha(x) = \|q_\alpha\|_{L^2}^{-2} \int_{\mathbb{R}^k} g(x, \psi(s)) q_\alpha(s) \mathrm{d}P_\xi(s).$$

Perhaps the most prevalent approach in the literature to determine the model's policy function is to compute g from a linearized version of the model. In this case

$$g(x;\vartheta) = A(\vartheta)x,$$

and *numeric* implementation of the methods proposed by Blanchard and Kahn (1980), Klein (2000) or Sims (2002) allows to solve for the matrix $A(\vartheta) \in \mathbb{R}^{n_x \times (n_x + n_y)}$ given any arbitrary but fixed $\vartheta \in \Theta$. Since the coefficients in the policy's PCE are here determined by

[6] Otherwise it is the orthogonal projection of h onto the subspace in $L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_\theta)$ spanned by multivariate polynomials in ψ^{-1} of total degree less or equal to N.

$$\hat{g}_\alpha(x) = \left(\|q_\alpha\|_{L^2}^{-2} \int_{\mathbb{R}^k} q_\alpha(s) A(\psi(s)) \mathrm{d}P_\xi(s) \right) x =: \hat{A}_\alpha x,$$

the series expansion of the linear policy function can be written as

$$g(x, \vartheta) = \sum_{\alpha \in \mathbb{N}_0^k} \hat{g}_\alpha(x) q_\alpha(\psi^{-1}(\vartheta)) = \left(\sum_{\alpha \in \mathbb{N}_0^k} \hat{A}_\alpha q_\alpha(\psi^{-1}(\vartheta)) \right) x.$$

Moreover, the $\hat{A}_\alpha$ coincides with the expansion coefficients from the PCE of the model outcome $A(\vartheta)$. Hence, the PCE of a linear policy is again linear and is represented by the polynomial expansion of the matrix-valued function $\vartheta \mapsto A(\vartheta)$.

A second popular approach to compute the model's policy function are projection methods.[7] In this approach g is constructed as a linear combination of some suitable basis functions Φ_i by

$$g(x;\vartheta) = \sum_{i=1}^{d} c_i(\vartheta) \Phi_i(x).$$

The coefficients in the PCE of g with respect to ϑ then satisfy

$$\hat{g}_\alpha(x) = \sum_{i=1}^{d} \left(\|q_\alpha\|_{L^2}^{-2} \int_{\mathbb{R}^k} q_\alpha(s) \big(c_i(\psi(s)) \big) \mathrm{d}P_\xi(s) \right) \Phi_i(x) =: \sum_{i=1}^{d} \hat{c}_{i\alpha} \Phi(x),$$

and the expansion of g can therefore be written as

$$g(x, \vartheta) = \sum_{\alpha \in \mathbb{N}_0^k} \hat{g}_\alpha(x) q_\alpha(\psi^{-1}(\vartheta)) = \sum_{i=1}^{d} \left(\sum_{\alpha \in \mathbb{N}_0^k} \hat{c}_{i\alpha} q_\alpha(\psi^{-1}(\vartheta)) \right) \Phi(x),$$

Now observe that the $\hat{c}_{i\alpha}$ coincide with the coefficients in the polynomial expansion of the model outcome $c_i(\vartheta)$, i.e., with the coefficients in the PCE of the coefficients of the projection solution. Consequently, the PCE of g is again a linear combination of the basis functions Φ_i and the coefficients are represented by the polynomial expansion of $\vartheta \mapsto c_i(\vartheta)$.

3 Numerical Analysis

This section presents the numerical implementation of a PCE for the benchmark RBC model. First, we analyze the convergence behavior of the series expansion for different model outcomes of interest. More specifically, the model outcomes include the solution, the second moments, and the impulse response functions from the model's

[7] See, for instance, Judd (1996), Chapter 11, Heer and Maußner (2024), Chapter 5, Judd (1992) or McGrattan (1999).

linear approximation. Additionally, we consider a global projection solution. Second, we compare different methods to compute the PCE coefficients regarding accuracy and efficiency. Lastly, we perform Monte-Carlo experiments, where we evaluate the performance of PCE for empirical applications as matching moments and likelihood-based approaches—both for linear and non-linear solutions.

3.1 The Model

We consider a benchmark RBC model where the social planner solves the following maximization problem

$$\begin{aligned}\max_{Y_t,C_t,N_t,K_{t+1}} U_0 :=& \mathbb{E}_0\left[\sum_{t=0}^{\infty}\beta^t\frac{C_t^{1-\eta}(1-N_t)^{\gamma(1-\eta)}}{1-\eta}\right],\\ \text{s.t. } C_t =& Y_t - K_{t+1} + (1-\delta)K_t,\\ Y_t =& e^{z_t}K_t^{\zeta}N_t^{1-\zeta},\\ \text{given } & K_0, z_0,\end{aligned}$$

where Y_t, C_t, N_t, and K_t denote output, consumption, working hours, and the capital stock, respectively. Moreover, the log of total factor productivity, z_t, evolves according to the AR(1) process

$$z_{t+1} = \rho z_t + \epsilon_{t+1},\ \epsilon_t \sim \text{iidN}(0,\sigma^2).$$

The predetermined state variables x_t and the non-predetermined control variables y_t are

$$x_t := \begin{pmatrix}K_t\\ z_t\end{pmatrix} \text{ and } y_t := \begin{pmatrix}Y_t\\ C_t\\ N_t\end{pmatrix}.$$

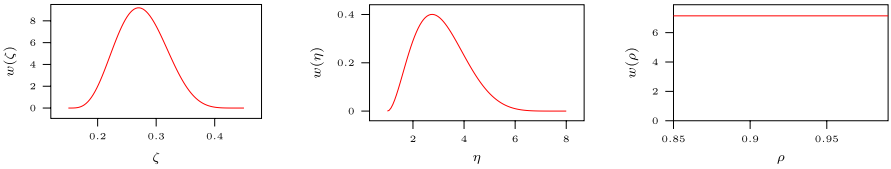
3.2 Convergence Behaviour

First, to study the basic convergence behavior of the PCE for various model outcomes in the benchmark RBC model, we consider an example where we set the uncertain parameters to $\theta := \begin{pmatrix}\zeta & \eta & \rho\end{pmatrix}$. Moreover, we assume the following probability distributions for the (stochastically independent) unknown parameters

$$\zeta \sim 0.15 + 0.3\cdot\text{Beta}(5,7),\ \eta \sim 1 + 7\cdot\text{Beta}(3,7),\ \rho \sim 0.85 + 0.14\cdot U(0,1).$$

The probability density functions with support $\Theta := [0.15;0.45]\times[1;8]\times[0.85;0.99]$ are illustrated in Fig. 1.

The transformations ψ_i between unknown parameters and germs are fixed as in Table 1 and the remaining parameters are calibrated as summarized in Table 2.



(a) Prior of ζ (b) Prior of η (c) Prior of ρ

Fig. 1 Distributions of uncertain parameters I

Table 2 Calibration I

Parameter	Description	Value
β	Discount factor	0.994
δ	Rate of capital depreciation	0.014
N	Steady state labor supply[1]	0.300
σ	Standard deviation	0.010

[1] Instead of pinning down the value of γ we set the steady state value of $N = 0.3$ and the model's steady state determines γ

Linear Policy Function

The first model outcome that we consider is the model's linear solution which is of the form

$$\begin{pmatrix} x_{t+1} \\ y_t \end{pmatrix} = A(\vartheta) x_t.$$

Given any parameter values $\vartheta \in \Theta$ the matrix $A(\vartheta) = \big(a_{ij}(\vartheta)\big)$, with $i = 1, \dots, 6$ and $j = 1, 2 \in \mathbb{R}^{6\times 2}$, can be easily computed numerically from available methods. As described in section 2.3, the expansion of the linear policy function is again linear and is represented by the polynomial expansion of $A(\vartheta)$. Hence, our task is to construct for each mapping $a_{ij} : \vartheta \mapsto a_{ij}(\vartheta)$ the truncated PCE[8]

$$a_{ij}^{(N)}(\vartheta) := S_N^{\text{tot}}(a_{ij} \circ \psi)(\psi^{-1}(\vartheta)) = \sum_{\alpha \in \mathbb{N}_0^3, |\alpha| \le N} \hat{a}_{ij\alpha} q_\alpha(\psi^{-1}(\vartheta)). \tag{10}$$

Moreover, we first want to abstract from errors in the computation of the expansion coefficients $\hat{a}_{ij\alpha}$ and to focus on the convergence behavior of $a_{ij}^{(N)} \to a_{ij}$ in L^2 as $N \to \infty$. Therefore, we compute the coefficients from full-grid Gauss-quadrature rules with a sufficiently large number of nodes which should guarantee that

[8] We only discuss the mappings $\vartheta \mapsto a_{ij}(\vartheta)$ for $i = 1, 3, \dots, 6$ and $j = 1, 2$ since the expansion of the exogenous AR(1)-process ($i = 2$) w.r.t. ρ is trivial.

integration errors in (5b) (where now $h = a_{ij}$) remain insignificant. More concretely, we apply $N + 5$ nodes in each of the three one-dimensional quadrature rules. We compute the coefficients from the quadrature rules and determine the L^2 error from

$$\begin{aligned} \|a_{ij}^{(N)} - a_{ij}\|_{L^2} &= \left(\int_{\mathbb{R}^3} \left(a_{ij}^{(N)}(\vartheta) - a_{ij}(\vartheta) \right)^2 dP_\theta \right)^{1/2} \\ &\approx \left(\frac{1}{M} \sum_{i=1}^{M} \left(a_{ij}^{(N)}(\vartheta^{(i)}) - a_{ij}(\vartheta^{(i)}) \right)^2 \right)^{1/2} \end{aligned} \tag{11}$$

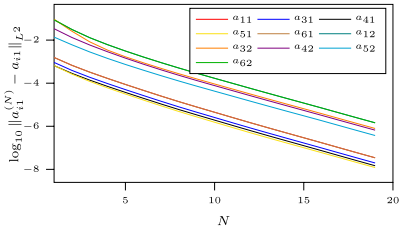
where we draw $M = 10^5$ iid sample points $\vartheta^{(i)}$ from the distribution of θ.

The results are presented in Fig. 2a in $\log_{10}$-base for $N = 1$ to $N = 19$ and suggest linear convergence of the series expansions for each a_{ij}. The L^2 error for all components of the matrix already falls to the order of magnitude of -3 for $N = 7$ and is as low as -6 for $N = 19$. Moreover, Fig. 2b also shows the time needed for all computations. In case of the PCE, the total time reported includes i) the computation of expansion coefficients $\hat{a}_{ij\alpha}$ from the full-grid quadrature rules which require $(N + 5)^3$ model evaluations and ii) the subsequent (trivial) evaluation of the truncated PCE $a_{ij}^{(N)}(\vartheta^{(i)})$ at the 100,000 sample points. For comparison, we also show the computational time that is required to determine the model solution $a_{ij}(\vartheta^{(i)})$ repeatedly at all 100,000 sample points. Most importantly, since even for $N = 19$ the number of model evaluations for the construction of the PCE is significantly smaller at 13824 than the number of evaluation points, the time required by the PCE remains less than one-third of the time needed for repeatedly solving the model.

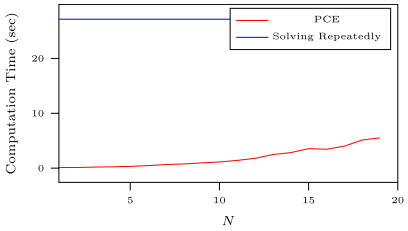
Second Moments

The second model outcomes we consider are the model's second moments. More specifically, we consider the variables' standard deviations and the correlations obtained from the model's linear policy. Instead of relying on simulations, we employ available formulae for moments of first-order autoregressive processes to the linear solution. We proceed the same way as in the preceding paragraph and compute for each moment, say x, a series expansion $x^{(N)} := \sum_{\alpha \in \mathbb{N}_0^3, |\alpha| \leq N} \hat{x}_\alpha q_\alpha(\psi^{-1}(\vartheta))$. Importantly, note that we directly construct the PCE of the second moments, i.e., of the mapping $\vartheta \mapsto x(\vartheta)$. An alternative approach to employ PCE for the second moments would be to first construct the PCE of the linear policy and subsequently use this PCE of the linear policy to compute the second moments.

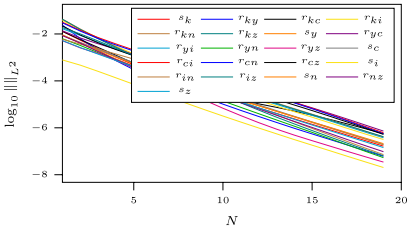
Figure 2c again shows linear convergence of the PCEs for each second moment. The L^2 error in the approximation of the model's moments has fallen to the order of magnitude of -3 by $N = 7$ and further declines to -6 by $N = 19$. Moreover, the computation time of the PCE versus the time for repeated computations of the model's moments is illustrated in Fig. 2d. For the same reasons as before, the time needed by the PCE remains throughout significantly lower than the time required for repeated calculations.



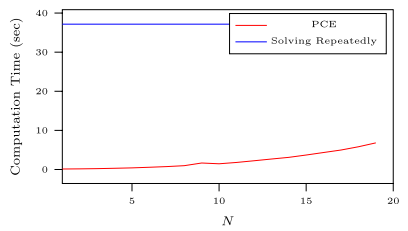
(a) L^2 convergence linear policy



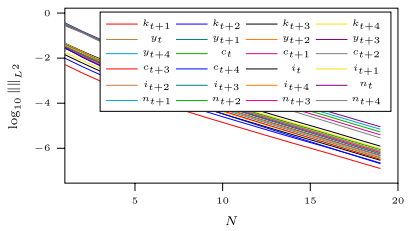
(b) Computation Time linear policy



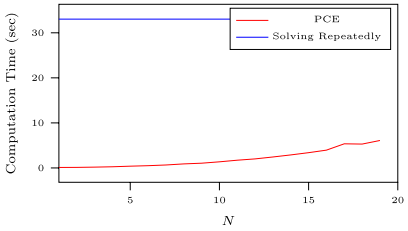
(c) L^2 convergence second moments



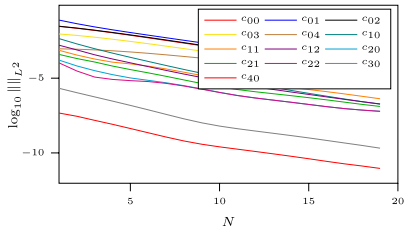
(d) Computation Time second moments



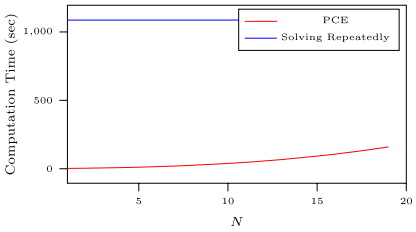
(e) L^2 convergence IRF



(f) Computation Time IRF



(g) L^2 convergence global solution



(h) Computation Time global solution

Fig. 2 L^2 convergence of PCE and computation time on an Intel® Core™i7-7700 CPU @ 3.60GHz

Impulse Response Function

The next model outcomes we discuss are the variables' impulse response functions in response to a one-time shock to TFP by one conditional standard deviation. For the sake of exposition, we only consider the variables' outcomes for the next four periods after the shock hits the economy and add the remark that the series

expansions become more trivial for later periods where the variables converge back to their stationary values. Hence, we construct PCEs for all variables' outcomes, say X_{t+s}, for periods $s = 0, \dots, 4$. Note again that the PCE is constructed directly for each mapping $\vartheta \mapsto X_{t+s}(\vartheta)$.

We show the L^2 errors over the unknown parameters' support in Fig. 2e. Convergence is again linear as $N \to \infty$ and the L^2 errors for all variables' outcomes fall to the order of magnitude of -5 by $N = 19$. Furthermore, the computation time of the PCE remains far below the time required for repeated computations of the model's IRFs.

Projection Solution

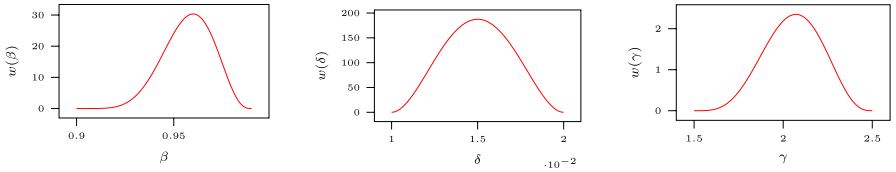
The last model outcome for which we want to illustrate the convergence behavior is the model's projection solution computed from Chebyshev polynomials as basis functions. More specifically, we define $k_t := \ln(K_t/K^\star(\vartheta))$ where $K^\star(\vartheta)$ is the capital stock's stationary solution and approximate the policy function for working hours by

$$n_t = g(k_t, z_t; \vartheta) = \sum_{i+j \leq 4} c_{i,j}(\vartheta) T_i\left(2\frac{k_t - k}{\bar{k} - k} - 1\right) T_j\left(2\frac{z_t - z}{\bar{z} - z} - 1\right),$$

where we further introduce the transformation $n_t := \ln(N_t/(1 - N_t))$. The T_i are Chebyshev polynomials of degree i and $[k;\bar{k}] \times [z;\bar{z}] = [\ln(0.8); -\ln(0.8)] \times [-3\frac{\sigma}{\sqrt{1-\rho^2}}; 3\frac{\sigma}{\sqrt{1-\rho^2}}]$ is the domain of the approximation g. The remaining variables are computed analytically from k_t, n_t and z_t and the coefficients $c_{i,j}(\vartheta)$ are determined in such a way that the model's Euler equation holds exactly at 13 appropriately selected collocation points.[9]

We discussed in section 2.3 that the expansion of the projection solution is again a linear combination of the same basis functions, i.e., of $T_{i_1} T_{i_2}$ with $i_1 + i_2 \leq 4$, and the coefficients are given by the series expansions of the mappings $\vartheta \mapsto c_{i,j}(\vartheta)$. Hence, we construct truncated PCEs, $c_{i,j}^{(N)} := \sum_{\alpha \in \mathbb{N}_0^3, |\alpha| \leq N} \hat{c}_{ij\alpha} q_\alpha(\psi^{-1}(\vartheta))$ from full-grid quadrature rules with $N + 5$ nodes in each dimension. The L^2 error, $\|c_{i,j}^{(N)} - c_{ij}\|_{L^2}$, in log10-basis is again decreasing linearly as $N \to \infty$ as displayed in Fig. 2g and the time for construction and evaluation of the PCEs in Fig. 2h remains throughout significantly smaller than the time for repeated computations of the global solution.

[9] The collocation points are combinations of the zeros of the Chebyshev polynomials in the approximation.



(a) Prior of β (b) Prior of δ (c) Prior of γ

Fig. 3 Distributions of uncertain parameters II

3.3 Computation of PCE Coefficients

In the previous subsection, our focus was on the convergence behavior of the PCE when the degree of truncation N was increased. We therefore abstracted from possible errors in the computation of the PCE coefficients and employed a full-grid quadrature rule with sufficiently many nodes. While full-grid quadrature rules have the favorable property that the number of nodes can be easily chosen in such a way that they provide exact integration rules for polynomials up to the desired degree, the number of nodes grows exponentially in the dimension of the parameter vector. Hence, they may provide the most convenient way for computation of the PCE coefficients when the number of unknown parameters is not too large, but they become quickly ineffective in higher dimensional problems. If the PCE coefficients are determined from alternative methods, the approximation error of the feasible PCE does not only include the error from truncation of the series expansion but also from a potentially less accurate approximation of the PCE coefficients that becomes necessary.

In this section, we now switch perspective and analyze the convergence behavior of the PCE when its coefficients are computed from different methods. Next to the benchmark full-grid quadrature rule, the PCE coefficients are additionally approximated by a sparse-grid Smolyak quadrature rule and by least squares.

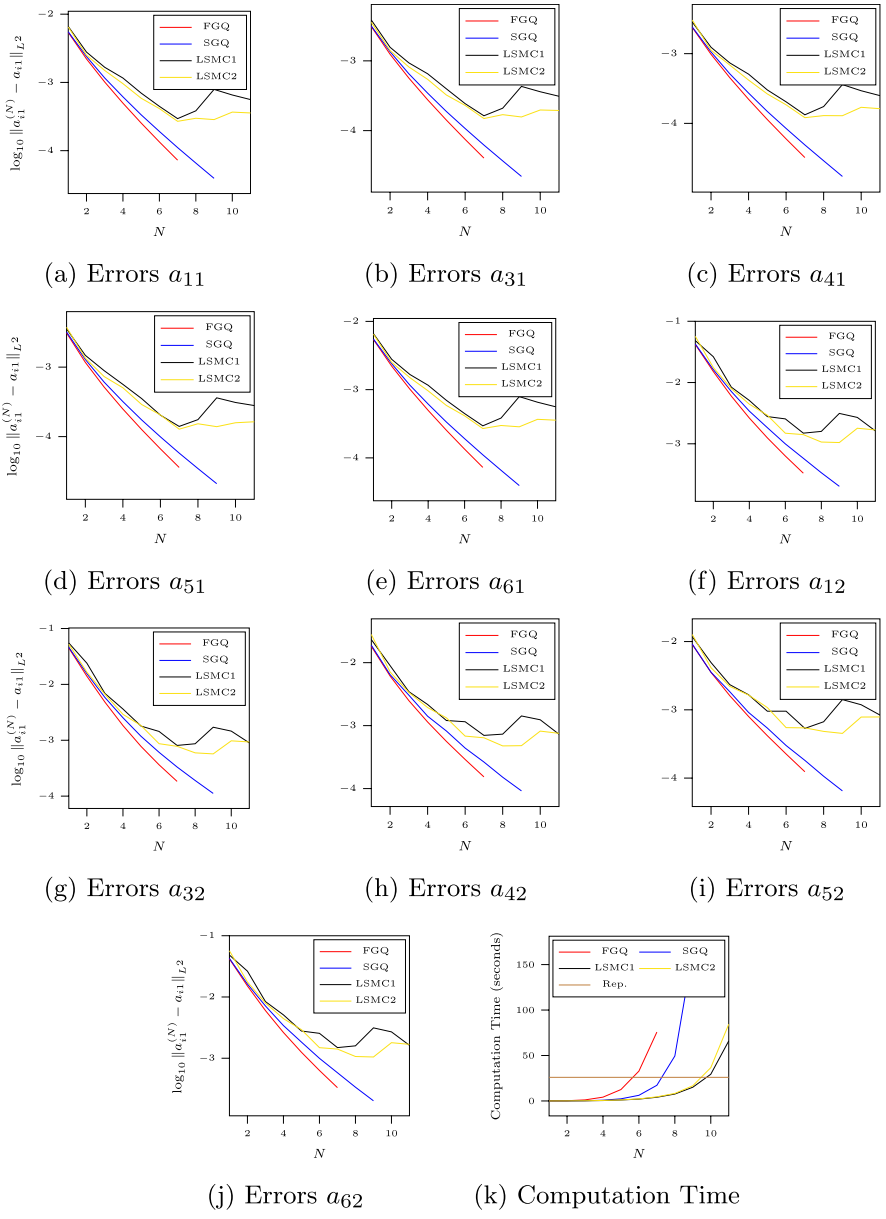
We apply our analysis to the PCE of the model's linear solution but now consider a higher dimensional problem. The vector of unknown parameters expands to $\theta := \begin{pmatrix}\zeta & \eta & \rho & \beta & \delta & \gamma\end{pmatrix}$.[10] The assumed distributions for ζ, η and ρ remain as before in Fig. 1, and the distributions of the additional unknown parameters are chosen as

$$\beta \sim 0.9 + 0.09 \cdot \text{Beta}(7,4),\ \delta \sim 0.01 + 0.01 \cdot \text{Beta}(3,3),\ \gamma \sim 1.5 + 1 \cdot \text{Beta}(5,4).$$

The probability densities for β, δ and γ are visualized in Fig. 3.

We compute the truncated PCE (10) for each mapping $a_{ij} : \vartheta \mapsto a_{ij}(\vartheta)$ in the linear policy $A(\vartheta) = \big(a_{ij}(\vartheta)\big)$, with $i = 1, \ldots, 6$ and $j = 1, 2 \in \mathbb{R}^{6\times 2}$. The PCE coefficients are now determined either by i) a full-grid Gauss quadrature rule with $N + 1$

[10] These are all of the model's parameters except the standard deviation σ which does not affect the model's linear policy.



(a) Errors a_{11} (b) Errors a_{31} (c) Errors a_{41}

(d) Errors a_{51} (e) Errors a_{61} (f) Errors a_{12}

(g) Errors a_{32} (h) Errors a_{42} (i) Errors a_{52}

(j) Errors a_{62} (k) Computation Time

Fig. 4 L^2 Convergence of PCE with approximated coefficients and computation time on an Intel® Core™i7-7700 CPU @ 3.60GHz I

nodes for each parameter (FGQ), ii) a sparse-grid Smolyak-Gauss quadrature rule with linear growth where the level is set in such a way that the one-dimensional quadrature rules include the nodes up to degree $N+1$ (SGQ), iii) least squares

where the number of sample point is set either twice (LSMC1) or iv) three times as large as the number of unknown PCE coefficients (LSMC2). After construction of the truncated PCE by each of the four methods, we compute the PCE's L^2 error as in (11) from a draw of $M = 10^5$ iid sample points from the parameter's distribution.

Figure 4 shows the convergence of the truncated (approximated) PCEs with approximated coefficients for increasing N. As expected, the PCE constructed from a full-grid quadrature rule, which should provide the most accurate determination of the coefficients, also shows the fastest convergence. It is followed by the PCE constructed from the sparse-grid Smolyak quadrature rule while the PCEs where the coefficients are computed by least squares perform worst. Since inaccuracies in the coefficients of higher degree polynomials may have large impact on the L^2 error of the PCE,[11] the PCEs computed from least squares even show increasing errors for larger N. Yet, the necessary computations for the full-grid quadrature method also require by far the most time. Figure 4k shows that by $N = 5$ the construction and evaluation of the PCE already consumes more time than 100,000 repeated computations of the model solution. In comparison, the sparse-grid quadrature rule is already significantly less computationally costly while the least-squares methods are least expensive to compute and remain less time-consuming than repeated computations of the model solution up to $N = 10$.

Finally, Fig. 5 provides a more convenient illustration of the different methods' efficiency and plots the PCEs' L^2 error versus the required computation time, both in $\log_{10}$-basis. According to this metric the full-grid quadrature method already performs worst and requires the most computation time to reach the same quality of approximation as the other methods. The most efficient method is the sparse-grid Smolyak quadrature rule. In the present case with six unknown parameters, it reaches an approximation with L^2 error of the order of magnitude of -4 before the required time for the PCE's construction exceeds the time for 100,000 repeated computations of the model solution.

3.4 Monte Carlo experiments for empirical methods

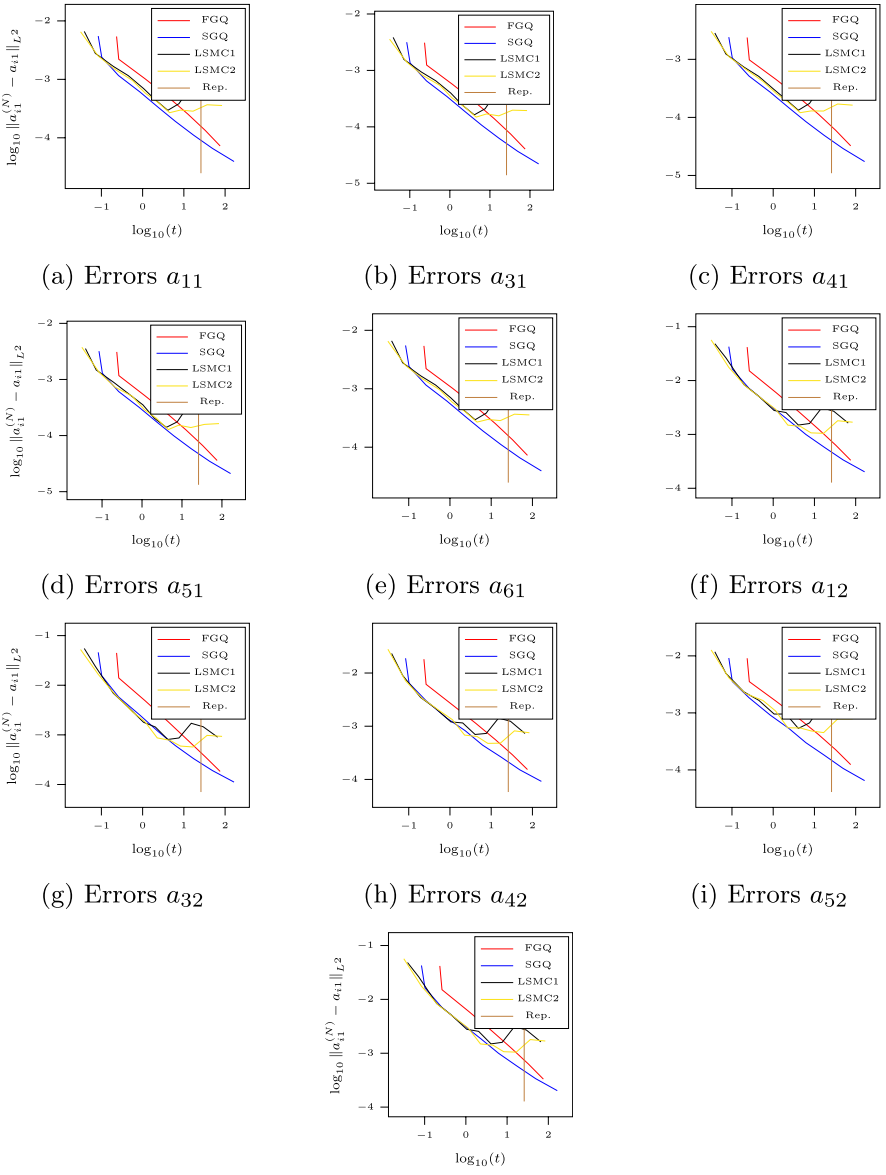
3.4.1 Estimation Based on Linearized Models

Design

Our Monte Carlo study for linearized models follows Ruge-Murcia (2007) and analyzes the performance of PCE when applied to different estimation methods. We set the vector of uncertain parameters to $\theta := (\beta, \rho, \sigma)$ and choose the following probability distributions with support $\Theta := [0.97;0.999] \times [0.75;0.995] \times [0.004;0.012]$ for the unknown parameters:

$$\beta \sim 0.97 + 0.029 \cdot \text{Beta}(2,2),\ \rho \sim 0.75 + 0.245 \cdot \text{Beta}(2,2),\ \sigma \sim 0.004 + 0.009 \cdot U(0,1).$$

[11] Note that the norm of the orthogonal polynomials, $\|q_\alpha\|_{L^2}$, is increasing in $|\alpha|$.



(a) Errors a_{11} (b) Errors a_{31} (c) Errors a_{41}

(d) Errors a_{51} (e) Errors a_{61} (f) Errors a_{12}

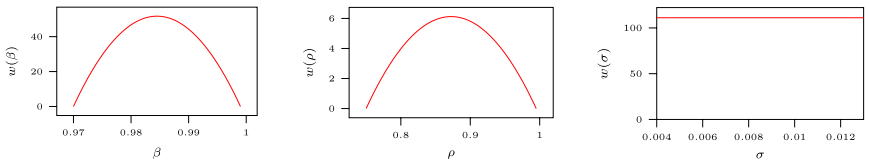
(g) Errors a_{32} (h) Errors a_{42} (i) Errors a_{52}

(j) Errors a_{62}

Fig. 5 L^2 Convergence of PCE with approximated coefficients and computation time on an Intel® Core™i7-7700 CPU @ 3.60GHz II

Figure 6 illustrates the uncertain parameters' probability densities and the remaining parameters are calibrated as summarized in Table 3.

The simulated data and the subsequent estimation of the parameters are both from a linearized model. While the advantage of PCE increases with more sophisticated



(a) Prior of β (b) Prior of ρ (c) Prior of σ

Fig. 6 Distributions of uncertain parameters III

Table 3 Calibration II

Fixed-parameter	Description	Value
ζ	Capital share	0.37
δ	Rate of capital depreciation	0.014
N	Steady-State labor supply	0.3
η	Risk aversion	2
Uncertain parameters	Description	Distribution
β	Discount factor	$\beta \sim 0.97 + 0.029 \cdot \text{Beta}(2,2)$
ρ	Persistence	$\rho \sim 0.75 + 0.245 \cdot \text{Beta}(2,2)$
σ	Standard deviation	$\sigma \sim 0.004 + 0.009 \cdot U(0,1)$

solution techniques, i.e., non-linear solutions, there is a reason to consider linear solutions: the existence of analytical representations of second moments and the likelihood function. Thus, differences in the estimated parameters between PCE and the benchmark (solving the model repeatedly) can be solely attributed to the approximation with PCE.

Matching Moments

To estimate the parameters by matching moments, we choose the following 5 targets: i) the variance of output and working hours, ii) the autocovariance (lag 1) of output and working hours, and iii) the covariance between output and working hours. We draw a sample $\vartheta^{(i)}, i = 1, \ldots, M$, of size $M = 1,000$ from the distribution of the unknown parameters. In a first step, we compute the linear approximation of the policy function and the second moments for each $\vartheta^{(i)}$ in the sample. Subsequently, we feed the computed second moments as targets to an optimizer and (point) estimate the unknown parameters by the method of matching moments. When minimizing the objective function, we distinguish the following three cases to evaluate the model's second moments for different parameter values: i) repeatedly solving the model and computing the second moments (benchmark), ii) constructing the PCE of the linear approximation of the policy function which we then evaluate and use to compute the variables' second moments ($h(\varphi) = g(x;\varphi)$) or iii) constructing the PCE of the model's second moments which we then evaluate ($h(\varphi)$ becomes directly the five mentioned

Table 4 Monte Carlo Results - GMM

Benchmark (repeated solution)			
Time	Total average 00:01.25		
PCE policy function			
Time	Total average 00:00.5	PCE 00:00.05	Estimation average 00:00.45
j	β	ρ	σ
$\bar{\epsilon}_j$	0.04	0.01	0.02
$\epsilon_{j,.05}$	0.00	0.00	0.00
$\epsilon_{j,.5}$	0.03	0.01	0.01
$\epsilon_{j,.95}$	0.11	0.03	0.06
PCE second moments			
Time	Total average 00:03.44	PCE 00:03.11	Estimation average 00:00.33
j	β	ρ	σ
$\bar{\epsilon}_j$	0.16	0.02	0.02
$\epsilon_{j,.05}$	0.02	0.00	0.00
$\epsilon_{j,.5}$	0.13	0.02	0.01
$\epsilon_{j,.95}$	0.43	0.06	0.09

Observable moments: variance of output, variance of hours, covariance between output and hours, autocovariance of output (lag 1), autocovariance of hours (lag 1). $\bar{\epsilon}_j$: mean error, $\epsilon_{j,.05}$: 5 percentile of error, $\epsilon_{j,.5}$: median of error, $\epsilon_{j,.95}$: 95 percentile of error. Errors of PCE-based methods are expressed as deviations from the benchmark method of repeatedly solving the policy function in percent of the range of the parameter's distribution. Time: mm:ss.f on an Intel® Core™i7-7700 CPU @ 3.60GHz. The truncation degree and quadrature level of the expanded policy function is 9 and of the second moments 19

targets). We compute the second moments either from analytic formulae for the linear solution (GMM) or from a simulation with $T = 10,000$ periods (SMM). We adapt the truncation degree and quadrature level manually to achieve sufficient accuracy to demonstrate the capabilities.[12] After obtaining the parameters' estimate $\hat{\vartheta}^{(i)}$, we define the PCE error by the deviation between the realized point estimate $\hat{\vartheta}^{(i)}_{\text{PCE}}$ from a PCE based method and the estimate $\hat{\vartheta}^{(i)}_{\text{BM}}$ obtained from the benchmark method, i.e.,

$$\epsilon_j^{(i)} = 100 \frac{\left|\hat{\vartheta}^{(i)}_{j,\text{PCE}} - \hat{\vartheta}^{(i)}_{j,\text{BM}}\right|}{\vartheta_{j,\max} - \vartheta_{j,\min}},\ j \in \{\beta, \rho, \sigma\},\ i = 1, ..., M,$$

where j indicates the estimator of the particular parameter and $\vartheta_{j,\max}$ and $\vartheta_{j,\min}$ denote the upper and lower bound of θ_j's prior support.

[12] We discuss heuristics for the choice of the truncation level below.

Table 5 Monte Carlo Results - SMM

Benchmark (repeated solution)			
Time	Total average 01:12.82		
PCE policy function			
Time	Total average 00:34.67	PCE 00:00.03	Estimation average 00:34.63
j	β	ρ	σ
$\bar{\epsilon}_j$	0.10	0.02	0.03
$\epsilon_{j,.05}$	0.01	0.00	0.00
$\epsilon_{j,.5}$	0.08	0.01	0.01
$\epsilon_{j,.95}$	0.25	0.04	0.16
PCE second moments			
Time	Total average 00:58.03	PCE 00:57.73	Estimation average 00:00.31
j	β	ρ	σ
$\bar{\epsilon}_j$	1.01	0.13	0.10
$\epsilon_{j,.05}$	0.10	0.01	0.01
$\epsilon_{j,.5}$	0.78	0.09	0.06
$\epsilon_{j,.95}$	2.62	0.35	0.30

Observable moments: variance of output, variance of hours, covariance between output and hours, autocovariance of output (lag 1), autocovariance of hours (lag 1). $\bar{\epsilon}_j$: mean error, $\epsilon_{j,.05}$: 5 percentile of error, $\epsilon_{j,.5}$: median of error, $\epsilon_{j,.95}$: 95 percentile of error. Errors of PCE-based methods are expressed as deviations from the benchmark method of repeatedly solving the policy function in percent of the range of the parameter's distribution. Time: mm:ss.f on an Intel® Core™i7-7700 CPU @ 3.60GHz. The truncation degree and quadrature level of the expanded policy function is 7 and of the second moments 13

Table 4 presents the results for GMM. We provide the computation time, the mean, the median, the 5 percentile, and the 95 percentile of the PCE error ϵ_j from $M = 1,000$ estimations. We find that the policy function's PCE provides a remarkably good approximation which results in deviations from the benchmark mostly smaller than one permille in comparison to the range of the parameter's distribution. The errors increase once PCE directly approximates the second moments. However, the average relative errors remain below two permille for all parameters and are almost always less than half a percent, again relative to the parameter's range. Using the PCE of the policy function reduces the computation time on average by 60 percent while the PCE of the second moments is more time-consuming than the benchmark. Nevertheless, the pure estimation procedure of the second moments' PCE is on average more than 25 percent faster than the estimation procedure of policy function's PCE.

Since analytic formulae for the model's moments are only available for the linear solution, GMM can only be employed for a linear approximation where computation time is rarely a limiting factor. If the model demands non-linear solutions, one has to resort to simulations to derive the model's moments. However, the computation

Table 6 Monte Carlo Results - Maximum Likelihood Estimation

Benchmark (repeated solution)			
Time	Total average 00:20.60		
PCE policy function			
Time	Total average 00:18.79	PCE 00:00.20	Estimation average 00:18.59
j	β	ρ	σ
$\bar{\epsilon}_j$	0.07	0.08	0.01
$\epsilon_{j,.05}$	0.00	0.00	0.00
$\epsilon_{j,.5}$	0.00	0.00	0.00
$\epsilon_{j,.95}$	0.00	0.01	0.02
PCE likelihood-function			
Time	Total average 00:10.81	PCE 00:10.36	Estimation average 00:00.44
j	β	ρ	σ
$\bar{\epsilon}_j$	0.08	0.40	0.05
$\epsilon_{j,.05}$	0.00	0.00	0.00
$\epsilon_{j,.5}$	0.00	0.08	0.01
$\epsilon_{j,.95}$	0.03	0.75	0.09

Observable: Output Y_t. $\bar{\epsilon}_j$: mean error, $\epsilon_{j,.05}$: 5 percentile of error, $\epsilon_{j,.5}$: median of error, $\epsilon_{j,.95}$: 95 percentile of error. Errors of PCE-based methods are expressed as deviations from the benchmark method of repeatedly solving the policy function in percent of the range of the parameter's distribution. Time: mm:ss.f on an Intel® Core™i7-7700 CPU @ 3.60GHz. The truncation degree and quadrature level of the expanded policy function is 9 and of the likelihood function 13

of non-linear solutions and the simulation of model outcomes increase the computational effort significantly. Working with the PCE of the policy function reduces the former burden while working with the PCE of the second moments helps to reduce both burdens. The results for our Monte-Carlo experiment with SMM are summarized in Table 5.

We find again that the policy function's PCE provides a remarkably good approximation which results in errors mostly smaller than 2.5 permille in comparison to the range of the parameter's distribution. Similar to GMM, errors rise if the model's second moments are directly approximated by PCE. However, the average relative errors remain around or below one percent for all parameters and are almost always less than 2.5 percent. Using the PCE of the policy function reduces the computation time on average by 50 percent while the PCE of the second moments reduces them only by 20 percent. However, the pure estimation procedure of the second moments' PCE is on average more than 99 percent faster than the estimation procedure of policy function's PCE. This illustrates the efficiency of expanding the QoI with PCE.

Table 7 Monte Carlo results - Bayesian estimation I

Benchmark (Repeated Solution)									
Time: Total average 08:36.11									
PCE Policy Function									
Time:		Total average 07:56.38			PCE 00:00.05		Estimation average 07:56.33		
		Mean:			Quantile:				
j	x:		5%	10%	25%	50%	75%	90%	95%
β	$\bar{\epsilon}_j(x)$	0.05	0.12	0.08	0.05	0.04	0.05	0.07	0.10
	$\epsilon_j(x)_{.05}$	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.01
	$\epsilon_j(x)_{.5}$	0.04	0.09	0.05	0.03	0.03	0.04	0.05	0.06
	$\epsilon_j(x)_{.95}$	0.15	0.33	0.23	0.15	0.15	0.14	0.22	0.33
ρ	$\bar{\epsilon}_j(x)$	0.23	0.32	0.28	0.25	0.25	0.30	0.37	0.45
	$\epsilon_j(x)_{.05}$	0.02	0.02	0.02	0.02	0.02	0.02	0.03	0.03
	$\epsilon_j(x)_{.5}$	0.19	0.24	0.24	0.20	0.21	0.25	0.30	0.35
	$\epsilon_j(x)_{.95}$	0.59	0.87	0.71	0.64	0.63	0.77	0.99	1.22
σ	$\bar{\epsilon}_j(x)$	0.09	0.11	0.10	0.09	0.09	0.11	0.15	0.20
	$\epsilon_j(x)_{.05}$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	$\epsilon_j(x)_{.5}$	0.07	0.09	0.08	0.07	0.07	0.09	0.12	0.15
	$\epsilon_j(x)_{.95}$	0.24	0.24	0.29	0.25	0.26	0.28	0.41	0.59

Observable: Output Y_t. $\bar{\epsilon}_j$: mean error, $\epsilon_{j,.05}$: 5 percentile of error, $\epsilon_{j,.5}$: median of error, $\epsilon_{j,.95}$: 95 percentile of error. Errors of PCE-based methods are expressed as deviations from the benchmark method of repeatedly solving the policy function in percent of the range of the parameter's distribution. Time: mm:ss.f on an Intel® Core™i7-7700 CPU @ 3.60GHz. The truncation degree and quadrature level of the expanded policy function is 9 and of the second moments 13

Likelihood-based Estimation

We proceed to analyze the performance of PCE in MLE and BE. More precisely, we now draw a sample of size $M = 500$ from the distribution of the unknown parameters. We approximate linearly the policy function and simulate a time series of output Y_t for $T = 200$ periods for each $\vartheta^{(i)}$ in the sample.[13] We treat the simulated time series as observations from which we either (point) estimate the parameters by MLE or conduct BE.

In the case of MLE, we distinguish the following three methods to evaluate the observations' likelihood for different parameter values: i) repeatedly solving the model and computing the likelihood (benchmark), ii) constructing the PCE of the linear approximation of the policy function which we then evaluate and use to compute the likelihood ($h(\varphi) = g(x;\varphi)$) or iii) constructing the PCE of the likelihood which we then evaluate ($h(\varphi) = L(Y_{1:T};\varphi)$). In order to avoid problems with weak identification and to focus on

[13] More precisely, we generate a sample of size $T = 300$ and burn the first 100 observations.

Table 8 Monte Carlo Results - Bayesian Estimation II

PCE Posterior-Kernel									
Time:		Total average 00:16.82			PCE 00:11.00		Estimation average 00:05.82		
		Mean:			Quantile:				
j	x:		5%	10%	25%	50%	75%	90%	95%
β	$\bar{\epsilon}_j(x)$	0.05	0.13	0.09	0.05	0.04	0.05	0.07	0.09
	$\epsilon_j(x)_{.05}$	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00
	$\epsilon_j(x)_{.5}$	0.04	0.08	0.06	0.03	0.03	0.03	0.04	0.06
	$\epsilon_j(x)_{.95}$	0.16	0.40	0.25	0.15	0.13	0.14	0.19	0.30
ρ	$\bar{\epsilon}_j(x)$	0.21	0.32	0.27	0.24	0.24	0.27	0.36	0.44
	$\epsilon_j(x)_{.05}$	0.02	0.02	0.02	0.01	0.02	0.02	0.02	0.03
	$\epsilon_j(x)_{.5}$	0.16	0.25	0.21	0.20	0.19	0.20	0.27	0.33
	$\epsilon_j(x)_{.95}$	0.59	0.86	0.69	0.59	0.60	0.72	1.01	1.19
σ	$\bar{\epsilon}_j(x)$	0.09	0.12	0.11	0.09	0.09	0.11	0.15	0.19
	$\epsilon_j(x)_{.05}$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	$\epsilon_j(x)_{.5}$	0.07	0.09	0.08	0.07	0.07	0.09	0.12	0.13
	$\epsilon_j(x)_{.95}$	0.24	0.32	0.30	0.25	0.24	0.29	0.42	0.55

Observable: Output Y_t. $\bar{\epsilon}_j$: mean error, $\epsilon_{j,.05}$: 5 percentile of error, $\epsilon_{j,.5}$: median of error, $\epsilon_{j,.95}$: 95 percentile of error. Errors of PCE-based methods are expressed as deviations from the benchmark method of repeatedly solving the policy function in percent of the range of the parameter's distribution. Time: mm:ss.f on an Intel® Core™i7-7700 CPU @ 3.60GHz. The truncation degree and quadrature level of the expanded policy function is 9 and of the second moments 13

the quality of PCE in the estimation procedure, here MLE is unusually applied to data in levels instead of the relative deviation from steady state.

For BE, the priors remain the same as in Table 3. Moreover, we again consider three methods to evaluate the posterior where the first two are analogous to i) and ii) above while iii) now involves constructing the PCE of the posterior's kernel ($h(\varphi) = L(Y_t;\varphi)p(\varphi)$, where $p(\varphi)$ is the prior of φ). For each of the three methods, we derive the posterior's mean as well as several quantiles of the posterior distribution from a standard random walk Metropolis Hasting (RWMH) algorithm with 100,000 draws from the posterior kernel.[14] We measure the accuracy of the PCE-based methods for each statistic of the posterior, say x, by computing the deviation between the statistic $\hat{x}^{(i)}_{j,\text{PCE}}$ obtained from the PCE based method and the statistic $\hat{x}^{(i)}_{j,\text{BM}}$ from the benchmark method by

$$\epsilon^{(i)}_{j,\text{PCE}}(x) = 100\frac{|\hat{x}^{(i)}_{j,\text{PCE}} - \hat{x}^{(i)}_{j,\text{BM}}|}{\vartheta_{j,\max} - \vartheta_{j,\min}}.$$

Again, we adapt the truncation degree and quadrature level manually to achieve sufficient accuracy.

[14] For the results we burn the first 50,000 draws.

Table 6 displays the results from MLE. First, deviations between the estimates from the method based on the policy function's PCE, the likelihood function's PCE, and the benchmark version remain remarkably small. The average error concerning the policy function's PCE estimation is smaller than one permille compared to the benchmark and relative to the range of the parameter. Furthermore, as the 95 percentile is smaller than the average, the error is mostly smaller than on average. The same holds for the estimation with the likelihood function's PCE. The average error is less than a half percent and the median is less than one permille. Using the PCE of the policy function does not reduce the computation time significantly, because the evaluation of the likelihood function is the time-consuming part. For this reason, using the PCE of the likelihood function is much more efficient. The total procedure is about 50 percent faster than the benchmark on average and the pure maximization procedure takes less than half a second on average.

Finally, Table 7 and Table 8 summarize the results from the PCE-based methods—approximation of the policy function or the kernel of the posterior—in BE. First, the errors between the two approximations are virtually the same. The average errors of the means and the medians are less than or equal to one-fourth of a percent. While deviations slightly increase for estimates of the posterior's lower and upper quantiles, they remain almost always less than 1.25 percent. Recognizing that errors may be partly caused by the RWMH algorithm itself, the deviations between the methods are negligible. Using the PCE of the policy function does not reduce the computation time significantly, because the evaluation of the likelihood function is likewise the time-consuming part. For this reason, the PCE of the likelihood function is much more efficient and nearly 99 percent faster than the benchmark.[15]

3.4.2 Estimation Based on the Global Solution

We proceed with our analysis by conducting the previous likelihood-based estimation for global, i.e., non-linear model solutions. On the one hand, the model's linear solution allowed an analytical derivation of the objective function of the estimations and, consequently, an exact assessment of the goodness of their PCE approximation. On the other hand, the solution and the derivation of the objective functions are fast by themselves. Consequently, time is not critical. Non-linear solutions and likelihood function evaluation with particle filters rely on numerical, partly Monte Carlo, methods, which makes the assessment vague. However, these methods are time-consuming, making PCE an interesting method to overcome these burdens.

We follow Fernández-Villaverde and Rubio-Ramírez (2005). The authors show that the non-linearities are crucial for parameter inference, even for our benchmark RBC model. We deviate from our previous study and follow Fernández-Villaverde and Rubio-Ramírez (2005) by considering only one true value for the parameters θ^o and the prior distribution choice, which is now uniform in all dimensions. The latter allows us to focus on the effects of the non-linear solution. The former is to evaluate

[15] It must be mentioned that a higher number of parameters leads to a decrease in efficiency.

our estimators by comparing the estimated average with the true values, as an exact objective function for the assessment is missing. We set the true parameter values $\theta^o = \{\beta^o, \rho^o, \omega^o\} = \{0.985, 0.9725, 0.0085\}$ and the priors

$$\beta \sim 0.98 + 0.01 \cdot U(0,1),\ \rho \sim 0.75 + 0.245 \cdot U(0,1),\ \sigma \sim 0.004 + 0.009 \cdot U(0,1).$$

Note that the domain of the priors for ρ and ω remains while for β, the domain shrinks. The latter is because β is well-identified. Outside this domain, the likelihood is too small ($< \exp(-1000)$) for an accurate particle filter evaluation. In the discussion below, we devote ourselves to cases where the model outcome is not well-defined or cannot be computed in a numerically stable way at all nodes of the quadrature rules.

Lastly, some information on the non-linear solution and the particle filter: we apply the projection solution described above with $[k;\bar{k}] \times [z;\bar{z}] = [\ln(0.9); -\ln(0.9)] \times [-2\frac{\sigma}{\sqrt{1-\rho^2}}; 2\frac{\sigma}{\sqrt{1-\rho^2}}]$ and use a generalized bootstrap particle filter with 2,000 particles (see Herbst & Schorfheide, 2016 Algorithm14). We conduct the exercises for $M = 96$ different datasets, each simulated using the globally solved model. If not otherwise stated, we still observe $T = 200$ periods of Y_t.

Maxmimum Likelihood

In the maximum likelihood analysis, we can only compare the maximum of the likelihood from the Kalman filter using a linear solution and of the PCE approximated likelihood as the likelihood directly from the particle filter is not differentiable—ruling out gradient-based optimizer. The literature refers to the use of differentiable likelihood surrogates or non-gradient-based optimizers. While the latter is a research topic itself, we contribute to the former idea by assessing the possibility of surrogate the likelihood with PCE.[16]

Figure 7 presents the results dependent on the truncation level ($N \in \{8, 9, ..., 14\}$). The upper three panels ((a)–(c)) display the bias of the estimators relative to the true parameter values, and the middle three panels ((d)–(f)) the relative standard deviations of the estimators. The last two panels ((g) and (f)) indicate the amount of a successful PCE approximation, i.e., inner maxima (g), and the time differences (f). It turns out that both approximations (linear solution and PCE surrogated likelihood) estimate on average β well. Both are on average within the range of $\pm 0.02\%$. The estimates for ρ are more biased. Yet, for truncations $N \geq 10$, the PCE estimator becomes noticeably less biased. The biggest difference between the estimation strategies is concerning σ. While for $N \geq 10$ the PCE estimates fluctuate close around the true value, the estimate from the linear solutions deviates on average by 3.25% from the parameter's true value. The analysis shows, that the estimators of the PCE surrogate are less or equal biased. Yet, the estimator's fluctuation is higher. However,

[16] Note that in our example the PCE likelihood surrogate MLE is on average more accurate than the average posterior modes from the repeated global solution sampler.

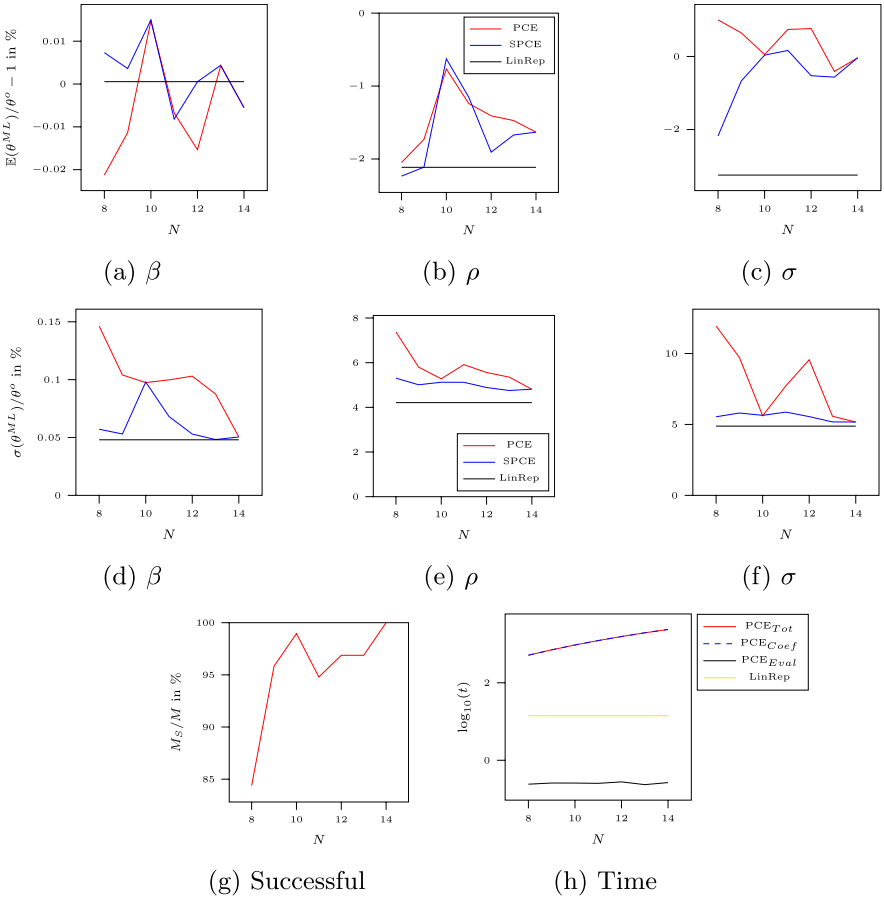
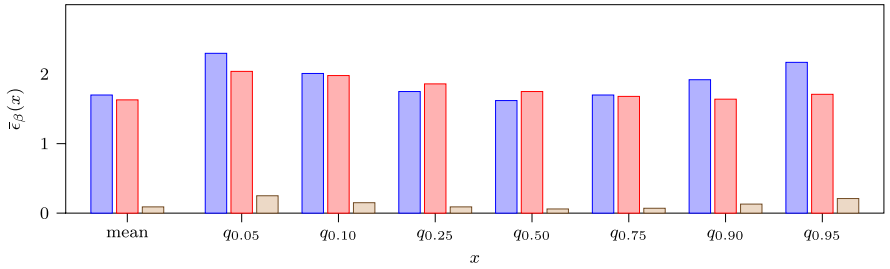


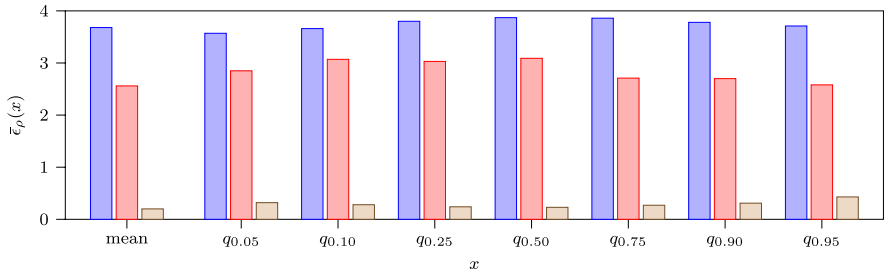
Fig. 7 ML from various likelihood approximations ($M = 96$). PCE: PCE approximated likelihood from a particle filter, SPCE: Only successful PCE approximations (M_S), i.e., exclusion of maxima at the parameter bounds. LinRep: Repeated likelihood evaluation using the Kalman Filter from the linear model solution. N equals the truncation level, the quadrature level equals N+1. Computation time on one core of an AMD® EPYC™7313 (Milan) CPU @ 3.00GHz

the estimator's standard deviation converges with N to the standard deviation of the linear solution estimates and is already similar for β and σ for $N \geq 13$.

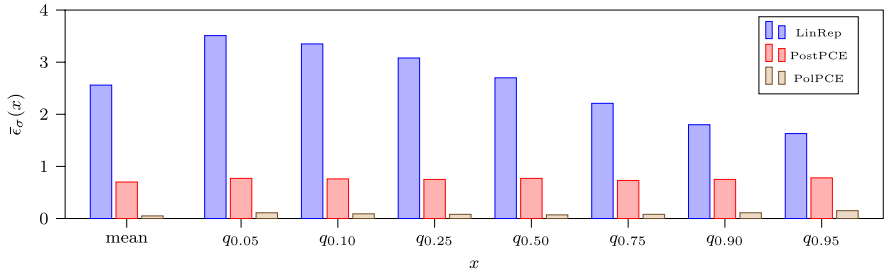
The amount of successful PCE, i.e., likelihood maxima at the bounds, increases from 85% for $N = 8$ above 95% for $N \geq 8$ and equals 100% for $N = 14$. One maximization with the PCE approximation takes on average between 9 min ($N = 8$) and 40 min ($N = 14$) and takes much longer than with the use of the linear solution (14



(a) β



(b) ρ



(c) σ

Fig. 8 Observable: Output Y_t. $\bar{\epsilon}_j$: mean error. Errors are expressed as deviations from the benchmark method of repeatedly solving the (global) policy function in percent of the range of the parameter's distribution. $N = 13$ equals the truncation level, the quadrature level equals $N + 1$. Performed on one core of an AMD® EPYC™7313 (Milan) CPU @ 3.00GHz

Table 9 Computational time comparison

	linear repeated	posterior PCE	policy fct. PCE	non-linear repeated
hh:mm:ss	00:05:45	00:32:27	17:45:11	18:17:37

$N = 13$ equals the truncation level, and the quadrature level equals $N + 1$. Performed on one core of an AMD® EPYC™7313 (Milan) CPU @ 3.00GHz

sec). However, the duration of the likelihood evaluation of the non-linear model is still quick and can be reduced easily and drastically via parallelization.[17]

Finally, the problem arises in whether the estimator's standard deviation and the remaining bias arise generally from the maximum likelihood method and the particle filter or from limitations of the PCE approximation. We can identify the reasons by improving the properties of the true MLE and the particle filter. To decrease the bias and standard deviation of the true MLE, we increase the number of observations (T=500) c.p., and to decrease the noise in the particle filter, we increase the number of particles to 10,000 c.p. Appendix 7 presents the results (Fig. 12 and 13). The additional information (T=500) leads to a similar decreasing standard deviation of both estimators the PCE surrogate likelihood and the likelihood from the model's linear approximation. However, while the bias of the MLE from the PCE surrogate likelihood shrinks further, the bias of the MLE from the linear solution only decreases for ρ. The bias for β and σ remain or even increase. Further, with more information, the PCE approximation becomes more stable. Regarding the higher amount of particles, there is unsurprisingly no improvement in the bias of the estimates from the PCE approximated likelihood. However, the estimator's standard deviation decreases for all considered truncation levels. With these two results, we conclude that PCE approximation errors are neither the drivers of the remaining inaccuracies nor limits a higher accuracy.

Bayesian Estimation

Note that in a Bayesian context besides the mode, we cannot observe the true statistics of the posterior distribution. Since the posterior statistics obtained from the global projection solution and the generic bootstrap particle filter should be at least unbiased (see Fernández-Villaverde & Rubio-Ramírez, 2005) we use them as a benchmark case and compare it with three other methods to evaluate the posterior: i) the linear approximate solution combined with a likelihood obtained from the Kalman-Filter, ii) the PCE surrogate of the posterior kernel, iii) and the PCE approximation of the global projection solution together with the likelihood from the generic particle filter. For both QoIs, we set the truncation and the quadrature level to $N = 13$ and $M = N + 1$, respectively.

As in the linear setup, we use the RWMH algorithm to generate 100,000 draws from the posterior distribution. However, since initializing the algorithm at the posterior mode is difficult when the likelihood is approximated by a particle filter (see the discussion above), we depart from the linear setup and specify the algorithm's proposal density using estimates of the posterior mean and variance. We obtain these estimates from 10,000 additional draws from a RWMH algorithm based on a proposal density pinned down by the prior's mean and variance.

Figure 8 displays the mean absolute deviations (relative to the range of the parameter's distribution) of the three competing methods to the benchmark case for

[17] We use here only one core. Hence the computational time for the PCE approximation can be roughly divided by the amount of available cores, e.g., with > 160 cores, the $N = 14$ approximation should become faster than the linear approximation, ignoring workers' allocation time.

various posterior statistics and Table 9 gives an overview over the average computational time for one estimation.[18]

For all three estimation parameters and all displayed statistics of the posterior, the PCE extension of the global projection solution yields the estimation results that come closest to the benchmark case. While the average absolute deviation is well below half a percent for all parameters, the average time required for one estimation (17h:45m:11s) is only around half an hour less compared to the benchmark (18h:17m:37s). However, this difference depends on the fraction of computational time of the solution on the total time. Note that the solution becomes quickly more time-consuming than the filter with an increasing number of states.

For β, the deviations between the PCE expansion of the posterior kernel and the benchmark case are similar to those of the linear approximation of the model solution. However, for ρ and in particular σ, the results are significantly closer to the benchmark method, with about one and almost two percentage points lower deviation. This, together with the fact that the computational time required (00h:32m:27s) is significantly lower, makes the PCE surrogate of the Posterior kernel a promising alternative to the benchmark method.

In contrast, the estimates based on the linear approximation of the model solution are computationally much more favorable (00h:05m:45s) but also deviate the most from the benchmark case with an average absolute deviation from just under two to almost four percent. In line with the results by Fernández-Villaverde and Rubio-Ramírez (2005), we document that for the parameter σ the deviations vary systematically for different percentiles of the posterior, as the mean absolute deviations between the linear and the global benchmark estimation method decrease by more than one and a half percentage points from the 5-th to the 95-th percentile.

Discussion

Our study of PCE for estimating a standard RBC model shows that the PCE-based methods deliver sufficient accurate surrogates to reproduce the results of estimates from the benchmark procedure—repeatedly solving the model. Gains in efficiency are larger than 50 percent for matching moments if the PCE of the policy function is used and for MLE if the PCE of the likelihood function is used. Additionally, the PCE of the likelihood from a particle filter is differentiable and, thus, enables a gradient-based optimization. Gains in efficiency are larger than 95 percent for BE with the chosen numbers of parameters, truncation degree, and quadrature level if the PCE of the posterior's kernel is used.

In our specification of the prior distributions we shape and shift the distributions to achieve compactness of the support. This procedure is unconventional in the Bayesian estimation of DSGE Models but helps for PCE. First and foremost, the compactness of the support helps to create a setting where the mapping from parameters to the model outcome is square-integrable. Second, it is indispensable for the construction of the PCE coefficients that the model outcome is well-defined and can

[18] We provide the complete estimation results (incl. various error percentiles) in Tables 12, 13 and 14 in Appendix 7.

be computed in a numerically stable way at all nodes of the quadrature rules.[19] Here, importance sampling for least squares, adaptive sparse grids, or grid domain reductions produces a remedy.[20]

In non-Bayesian approaches, the application of PCE demands the otherwise unnecessary specification of prior distributions. As L^2 convergence of the series expansion is achieved w.r.t. this prior distribution of the parameters, estimation results become less accurate if the true parameter value is at odds with the choice of priors, especially if the true parameter is outside the prior domain.[21]

Similarly, Lu et al. (2015) show that using PCE for BE may be inaccurate in two cases. First, the QoI is represented poorly by a low-order polynomial. Second, the posterior mass is in other regions than the prior mass. To solve these problems, they suggest an adaptive increasing polynomial order by verifying the accuracy at the next evaluation point. As our manual adaption is usually not feasible as it requires the benchmark results, this is also a practical method for determining the truncation level in general. In addition, a small magnitude of the Nth Fourier coefficient indicates a sufficiently high truncation level.

As PCE is a spectral decomposition approximated with a truncated polynomial expansion, generally, Runge's and Gibbs' phenomena could arise. Both result in spurious oscillation. Yet, using Gaussian quadratures and nodes prevents the former phenomenon, and the latter phenomenon only appears in the presence of discontinuity jumps. Problems with the approximation of a flat function are unknown. Thus, the frequent lack of identification of DSGE models does not challenge PCE itself.

Concerning time, the success of PCE is determined by the ratio of the number of model evaluations necessary to compute the coefficients and the number of model evaluations for the exercise at hand. Hence, PCE works best in cases with a small number of unknown parameters where the exercise demands many model evaluations. On the one hand, PCE loses efficiency in higher dimensional problems. On the other hand, most exercises are recursive (Monte Carlo sampler, gradient-based optimizer, etc.), where the model evaluations are independent of each other for constructing PCE. This independence makes the costly evaluations parallelizable—reducing the curse of dimensionality drastically with cluster or cloud computing. In addition, Soize and Desceliers (2010) develop tools to reduce the evaluation time of the constructed PCE.

Finally, our analysis is limited to an ergodic, stable stochastic process. However, Ozen and Bal (2016) show that, with some adaptions, PCE becomes suitable for time-dependent solutions and Jacquelin et al. (2015) for models with deterministic eigenfrequencies.

[19] For example, larger values of the capital share quickly result in numerical problems for the computation of the linear approximation of the policy function, and too large distances to the true parameter result in minus infinity log-likelihood values.

[20] For the latter, note that the priors must not change as otherwise information from the data would enter the prior.

[21] To put it simply, the prior distribution in such cases is only a guess that determines the accuracy of the solution in different ranges of parameter values.

4 Conclusion

The present article discusses the suitability of PCE for computational models in economics. For this purpose, we first provide the theoretical framework for PCE, review the basic theory, and give an overview of common distributions and corresponding orthogonal polynomials. We show how to use the expansion as a pointwise approximation for the QoI, e.g., to surrogate the linearized policy function or a policy function based on projection methods.

Second, we analyze PCE when applied to a standard RBC model and provide practical insights. We study convergence behavior for various QoIs and compare the most common methods to compute the PCE coefficients for a lower dimensional and a higher dimensional problem. Monte Carlo experiments for different empirical methods show that the PCE-based methods can accurately reproduce the results of the benchmark method of repeatedly solving the model. Gains in efficiency are large, especially for Bayesian inference.

Our discussion addresses potential drawbacks of the method. First, the efficiency of PCE suffers from the curse of dimensions in problems with numerous unknown parameters. Further, poorly chosen priors may affect the accuracy of the estimates.

PCE is a powerful tool for a broad set of applications and the recent literature addresses the highlighted drawback. We hope this article can encourage applications of PCE in economics. Especially, for parameter inference in complex models where numerous repeated solutions are infeasible or when time is critical as in real-time analysis of high-frequency data.

5 Supplementary information

MATLAB® code and replication file are available at www.johanneshuber.de/PCE.

Appendix 1 A Simple Example

Here, we want to outline the concept at a simple but analytically tractable construction of a PCE. Since our numerical analysis focuses on discretely-timed models, our example considers the following system of linear first-order difference equations in two real-valued variables $x_{1,t}$ and $x_{2,t}$,

$$\begin{aligned} \vartheta x_{1,t+1} + x_{2,t+1} &= x_{1,t}, \\ x_{1,t+1} + x_{2,t+1} &= x_{2,t}, \end{aligned}$$

for all $t \in \mathbb{N}$, and given $x_{1,0}$ and $x_{2,0}$. Moreover, $\vartheta \in (0, 1)$ is an unknown parameter. While the variables' explicit recursion can be derived straightforwardly here by

$$\begin{pmatrix} x_{1,t+1} \\ x_{2,t+1} \end{pmatrix} = H(\vartheta) \begin{pmatrix} x_{1,t} \\ x_{2,t} \end{pmatrix}, \text{ where } H(\vartheta) := \begin{pmatrix} h_{11}(\vartheta) & h_{12}(\vartheta) \\ h_{21}(\vartheta) & h_{22}(\vartheta) \end{pmatrix} = \begin{pmatrix} \frac{-1}{1-\vartheta} & \frac{1}{1-\vartheta} \\ \frac{1}{1-\vartheta} & \frac{-\vartheta}{1-\vartheta} \end{pmatrix},$$

the mapping $\vartheta \mapsto H(\vartheta)$ from the unknown parameter to the (linearized) policy can typically not be derived analytically, but can only be computed numerically if the system of difference equations is non-linear and stochastic. In consequence, if $H(\vartheta)$

needs to be computed for different parameter values, the underlying numerical methods must eventually be applied repeatedly. PCE, on the other hand, aims to represent the mapping $\vartheta \mapsto H(\vartheta)$ as a truncation from the Fourier series

$$h_{ij}(\vartheta) = \sum_{n=0}^{\infty} \hat{h}_{ij}^{(n)} q_n(\psi^{-1}(\vartheta)),$$

where q_n is the n-th polynomial from a family of orthogonal polynomials, $\psi^{-1}(\vartheta)$ is a transformation of the parameter space into the space of the polynomial orthogonal counterpart's argument, and $\hat{h}_{ij}^{(n)}$ is the corresponding Fourier coefficient of the polynomial. The truncated series expansion is constructed from a limited number of numerical evaluations of the mapping as follows.

First, the uncertainty about the parameter is taken into account by describing it by a random variable θ with suitable probability distribution P_θ. For the present example, suppose that θ is uniformly distributed over the interval $(0, b)$, $0 < b \leq 1$. Second, the series expansion is constructed in a well-known family of orthogonal polynomials, which satisfies orthogonality w.r.t. some weighting function w. Thereby, the appropriate family of orthogonal polynomials is most conveniently chosen in such a way that the weighting function w coincides with the probability density function of the unknown parameter. However, in order to achieve conformity between the weighting function and the density function, a (linear) transformation of the parameter typically becomes necessary. In the present case, Legendre polynomials $\{L_n\}_{n\geq 0}$ are orthogonal w.r.t. the weighting function $w(s) = \mathbb{1}_{(-1,1)}(s)$, i.e., they satisfy

$$\int_{\mathbb{R}} L_n(s)L_m(s)w(s)\mathrm{d}s = \begin{cases} 0, & \text{if } n \neq m, \\ \|L_n\|^2 := \frac{2}{2n+1}, & \text{if } n = m. \end{cases}$$

Hence, transformation of the unknown parameter θ to the so-called germ ξ by

$$\xi := \psi^{-1}(\theta) := 2\frac{\theta}{b} - 1 \;\Leftrightarrow\; \theta = \psi(\xi) = \frac{(\xi + 1)b}{2},$$

yields the desired result, and Legendre polynomials are orthogonal w.r.t. the probability distribution P_ξ of ξ. Given that $b < 1$, the mapping $s \mapsto h_{ij}(\psi(s))$ for each entry h_{ij} of the matrix H is square integrable w.r.t. P_ξ and can be represented by a Fourier series of the form[22]

$$h_{ij}(\psi(s)) = \sum_{n=0}^{\infty} \hat{h}_{ij}^{(n)} L_n(s). \tag{12}$$

Moreover, orthogonality implies that the Fourier coefficients $\hat{h}_{ij}^{(n)}$ satisfy

[22] The details in which sense convergence of the series can be established are discussed in the next section.

Table 10 Example

i	ω_i	s_i	ϑ_i	$h_{11}(\vartheta_i)$
1	0.2369	−0.9062	0.0422	−1.0441
2	0.4786	−0.5385	0.2077	−1.2621
3	0.5689	0	0.4500	−1.8182
4	0.4786	0.5385	0.6923	−3.2500
5	0.2369	0.9062	0.8578	−7.0314

$$\hat{h}_{ij}^{(n)} = \|L_n\|^{-2} \int_{-1}^{1} h_{ij}(\psi(s))L_n(s)\mathrm{d}s.$$

Finally, numerical integration methods are generally required to compute the coefficients $\hat{h}_{ij}^{(n)}$. For example, using Gauss-Legendre-quadrature with M nodes s_i and weights ω_i yields[23]

$$\hat{h}_{ij}^{(n)} \approx \|L_n\|^{-2} \sum_{i=1}^{M} h_{ij}(\psi(s_i))L_n(s_i)\omega_i.$$

Table 10 shows for $b = 0.9$ and $M = 5$ the quadrature weights ω_i, the nodes s_i, the corresponding retransformed parameter values $\vartheta_i := \psi(s_i)$, and for the matrix entry h_{11} the evaluation $h_{11}(\vartheta_i) = \frac{-1}{1-\vartheta_i}$.

Together with $L_0(s_i) = 1$, $L_1(s_i) = s_i$, $\|L_0\|^2 = 2$, and $\|L_1\|^2 = \frac{2}{3}$, one can therefore compute, e.g.,[24]

[23] If we additionally write the transformation ψ between parameter and germ in terms of the Legendre polynomials, i.e.,

$$\psi(s) = \underbrace{\frac{b}{2}}_{=:\hat{\vartheta}_0} L_0(s) + \underbrace{\frac{b}{2}}_{=:\hat{\vartheta}_0} L_1(s),$$

we equivalently arrive at

$$\hat{h}_{ij}^{(n)} \approx \|L_n\|^{-2} \sum_{i=1}^{M} h_{ij}\big(\hat{\vartheta}_0 L_0(s_i) + \hat{\vartheta}_1 L_1(s_i)\big)L_n(s_i)\omega_i.$$

Note that this expression is identical to the more general form in (3).

[24] For comparison, exact integration yields

$$\hat{h}_{11}^{(0)} = \frac{1}{2}\int_{-1}^{1} \frac{-1}{1-\frac{(s+1)b}{2}}\mathrm{d}s = \frac{\ln(1-b)}{b} = -2.56,$$

$$\hat{h}_{11}^{(1)} = \frac{3}{2}\int_{-1}^{1} \frac{-s}{1-\frac{(s+1)b}{2}}\mathrm{d}s = \frac{6-3b}{b^2}\ln(1-b) + \frac{6}{b} = -2.71.$$

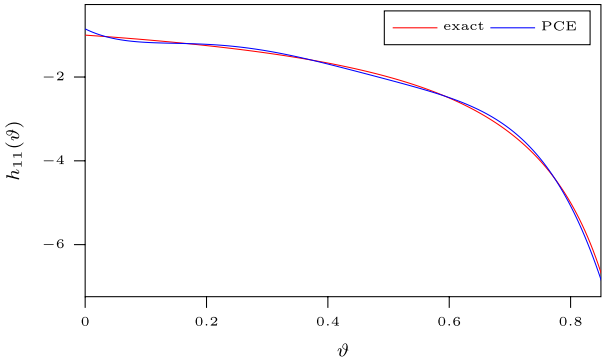


Fig. 9 Example: Exact Evaluation and PCE (numerical integration)

$$\hat{h}_{11}^{(0)} \approx \frac{1}{2}\sum_{i=1}^{5} h_{11}(\vartheta_i)\omega_i = -2.55 \quad \text{and} \quad \hat{h}_{11}^{(1)} \approx \frac{3}{2}\sum_{i=1}^{5} h_{11}(\vartheta_i)s_i\omega_i = -2.70.$$

In this case, the computation of the Fourier coefficients $\hat{h}_{11}^{(n)}$ requires $M = 5$ (numerical) evaluations of the mapping $\vartheta \mapsto h_{11}(\vartheta)$. After computation of the first $N+1$ Fourier coefficients, one can use the truncated series expansion of (12), i.e.,

$$h_{11}(\vartheta) \approx \sum_{n=0}^{N} \hat{h}_{11}^{(n)} L_n(\psi^{-1}(\vartheta)),$$

to (approximately) evaluate $h_{11}(\vartheta)$ for arbitrary parameter values without further need for direct numerical evaluations.[25] Figure 9 shows a comparison between the exact evaluation of $h_{11}(\vartheta)$ and the truncated PCE with truncation level $N = 5$.

Finally, note already here that an important restriction of the methods is the requirement that the mapping $s \mapsto h_{ij}(\psi(s))$ is square integrable w.r.t. P_ξ, or equivalently w.r.t. the weighting function w corresponding to the family of orthogonal polynomials. In the present example, this condition is fulfilled for $b < 1$. Yet, if $b = 1$, the integrals from which the coefficients are defined are not finite, e.g.,

$$\hat{h}_{11}^{(0)} = \frac{1}{2}\int_{-1}^{1} \frac{-1}{1-\frac{s+1}{2}} \mathrm{d}s = -\infty.$$

[25] Of course, an appropriate choice of the number M of quadrature nodes and, therefore, of the number of numerical evaluations is necessary to derive the Fourier coefficients depending on the truncation level N. More details on this topic are provided in the next section.

Appendix 2 Orthogonal Polynomials

We give a short overview of the families of orthogonal polynomials summarized in Table 1. More details, in particular regarding their completeness in the respective Hilbert spaces L^2 of square-integrable functions, can be found in Szegő (1939).

Hermite polynomials

Hermite polynomials are defined by the recurrence relation

$$H_0(x) = 1,\ H_1(x) = 2x,\ H_{n+1}(x) = 2xP_n(x) - 2nP_{n-1}(x), n \geq 2$$

and form a complete orthogonal system on $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \tilde{w}(x)\mathrm{d}x)$ with weighting function

$$\tilde{w}(x) := e^{-x^2}.$$

More specifically,

$$\int_{\mathbb{R}} H_n(x)H_m(x)\tilde{w}(x)\mathrm{d}x = 2^n(n!)\sqrt{\pi}\delta_{n,m}$$

The probability density function of a normal distributed random variable $\theta \sim \mathrm{N}(\mu, \sigma^2)$ with mean μ and variance σ^2 is given by

$$f_\theta(\vartheta) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\vartheta-\mu)^2}{2\sigma^2}}.$$

Fixing the transformation between the germ and θ in this case to

$$\psi(s) := \mu + \sqrt{2}\sigma s$$

so that the germ ξ is defined by

$$\xi := \psi^{-1}(\theta) = \frac{\theta - \mu}{\sqrt{2}\sigma}$$

implies that ξ has probability density function

$$w(s) = f_\theta(\psi(s))\psi'(s) = \frac{1}{\sqrt{\pi}} e^{-s^2} = \frac{1}{\sqrt{\pi}}\tilde{w}(s).$$

Since w differs from $\tilde{w}$ only by a constant factor, it follows that

$$L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{d}P_\xi) = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), w(s)\mathrm{d}s) = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \tilde{w}(s)\mathrm{d}s),$$

and that Hermite polynomials also form a complete orthogonal system in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{d}P_\xi)$ with

$$\begin{aligned}\int_{\mathbb{R}} H_n(s)H_m(s)\mathrm{d}P_\xi(s) &= \int_{\mathbb{R}} H_n(s)H_m(s)w(s)\mathrm{d}s \\ &= \frac{1}{\sqrt{\pi}}\int_{\mathbb{R}} H_n(s)H_m(s)\tilde{w}(s)\mathrm{d}s \\ &= 2^n(n!)\delta_{n,m}.\end{aligned}$$

Moreover, given the nodes s_j and weights $\tilde{\omega}_j$ from the common Gauss-Hermite-quadrature rule for weighting function $\tilde{w}$, the Gauss-quadrature rule in terms of weighting function w has the same nodes while the weights are scaled by $\omega_j = \frac{\tilde{\omega}_j}{\sqrt{\pi}}$.

Legendre polynomials

Legendre polynomials are defined by the recurrence relation

$$L_0(x) = 1, \quad L_1(x) = 2x, \quad (n+1)L_{n+1}(x) = (2n+1)xL_n(x) - nL_{n-1}(x),\ n \geq 2$$

and form a complete orthogonal system in $L^2([-1,1], \mathcal{B}([-1,1]), \mathrm{d}x)$, i.e.,

$$\int_{-1}^{1} L_n(x)L_m(x)\mathrm{d}x = \frac{2}{2n+1}\delta_{n,m}.$$

The probability density function of an uniformly distributed random variable $\theta \sim \mathrm{U}[0,1]$ over $[0,1]$ is given by

$$f_\theta(\vartheta) = \mathbb{1}_{[0,1]}(\vartheta) := \begin{cases} 1, & \text{if } \vartheta \in [0,1] \\ 0, & \text{if } \vartheta \in \mathbb{R} \setminus [0,1] \end{cases}$$

Fixing the transformation between the germ and θ in this case to

$$\psi(s) := \frac{s+1}{2}$$

so that the germ ξ is defined by

$$\xi := \psi^{-1}(\theta) = 2\theta - 1$$

implies that ξ has probability density function

$$w(s) = f_\theta(\psi(s))\psi'(s) = \frac{1}{2}\mathbb{1}_{[-1,1]}(s).$$

Hence, it follows that

$$L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{d}P_\xi) = L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), w(s)\mathrm{d}s) \simeq L^2([-1,1], \mathcal{B}([-1,1]), \mathrm{d}s),$$

and consequently the Legendre polynomials also form a complete orthogonal system in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{d}P_\xi)$ with

$$\begin{aligned}\int_{\mathbb{R}} L_n(s)L_m(s)\mathrm{d}P_\xi(s) &= \int_{\mathbb{R}} L_n(s)L_m(s)w(s)\mathrm{d}s \\ &= \frac{1}{2}\int_{-1}^{1} L_n(s)L_m(s)\mathrm{d}s \\ &= \frac{1}{2n+1}\delta_{n,m}.\end{aligned}$$

Moreover, given the nodes s_j and weights $\tilde{\omega}_j$ from the common Gauss-Legendre quadrature rule for weighting function $\tilde{w}$, the Gauss-quadrature rule in terms of weighting function w has the same nodes while the weights are scaled by $\omega_j = \frac{\tilde{\omega}_j}{2}$.

Jacobi polynomials

Jacobi polynomials are defined by the recurrence relation

$$\begin{aligned}&J_0^{(\alpha,\beta)}(x) = 1, \\ &J_1^{(\alpha,\beta)}(x) = \frac{1}{2}(\alpha-\beta+(\alpha+\beta+2)x), \\ &a_{1,n}J_{n+1}^{(\alpha,\beta)}(x) = (a_{2,n}+a_{3,n}x)J_n^{(\alpha,\beta)}(x) - a_{4,n}J_{n-1}^{(\alpha,\beta)}(x), n \geq 2\end{aligned}$$

where

$$\begin{aligned}a_{1,n} &= 2(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta), \\ a_{2,n} &= (2n+\alpha+\beta+1)(\alpha^2-\beta^2), \\ a_{3,n} &= (2n+\alpha+\beta)(2n+\alpha+\beta+1)(2n+\alpha+\beta+2), \\ a_{4,n} &= 2(n+\alpha)(n+\beta)(2n+\alpha+\beta+2).\end{aligned}$$

They form a complete orthogonal system on $L^2([-1,1],\mathcal{B}([-1,1]),\tilde{w}(x)\mathrm{d}x)$ with weighting function

$$\tilde{w}(x;\alpha,\beta) := (1-x)^\alpha(1+x)^\beta.$$

More specifically,

$$\begin{aligned}&\int_{-1}^{1} J_n^{(\alpha,\beta)}(x)J_m^{(\alpha,\beta)}(x)\tilde{w}(x;\alpha,\beta)\mathrm{d}x \\ &\qquad = \frac{2^{\alpha+\beta+1}}{2n+\alpha+\beta+1}\frac{\Gamma(n+\alpha+1)\Gamma(n+\beta+1)}{\Gamma(n+\alpha+\beta+1)n!}\delta_{nm}.\end{aligned}$$

The probability density function of a Beta-distributed random variable $\theta \sim \mathrm{Beta}(\alpha,\beta)$ with shape parameters α and β is given by [26]

[26] We denote by $B(x, y)$ the beta function.

$$f_\theta(\vartheta;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\vartheta^{\alpha-1}(1-\vartheta)^{\beta-1}\mathbb{1}_{[0,1]}(\vartheta).$$

Fixing the transformation between the germ and θ in this case to

$$\psi(s) := \frac{s+1}{2}$$

so that the germ ξ is defined by

$$\xi := \psi^{-1}(\theta) = 2\theta - 1$$

implies that ξ has probability density function

$$\begin{aligned} w(s;\alpha,\beta) &= f_\theta(\psi(s);\alpha,\beta)\psi'(s) \\ &= \frac{1}{B(\alpha,\beta)}\left(\frac{s+1}{2}\right)^{\alpha-1}\left(1-\frac{s+1}{2}\right)^{\beta-1}\frac{1}{2}\mathbb{1}_{[-1,1]}(s) \\ &= \frac{2^{1-\alpha-\beta}}{B(\alpha,\beta)}(s+1)^{\alpha-1}(1-s)^{\beta-1}\mathbb{1}_{[-1,1]}(s) \\ &= \frac{2^{1-\alpha-\beta}}{B(\alpha,\beta)}\tilde{w}(s;\beta-1,\alpha-1)\mathbb{1}_{[-1,1]}(s). \end{aligned}$$

Since $w(s;\alpha,\beta)$ differs from $\tilde{w}(s;\beta-1,\alpha-1)$ only by a constant factor, it follows that

$$\begin{aligned} L^2(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathrm{d}P_\xi) = L^2(\mathbb{R},\mathcal{B}(\mathbb{R}),w(s;\alpha,\beta)\mathrm{d}s) \simeq \\ \simeq L^2([-1,1],\mathcal{B}([-1,1]),\tilde{w}(s;\beta-1,\alpha-1)\mathrm{d}s), \end{aligned}$$

and that the Jacobi polynomials $\{J_n^{(\beta-1,\alpha-1)}\}_{n\in\mathbb{N}_0}$ also form a complete orthogonal system in $L^2(\mathbb{R},\mathcal{B}(\mathbb{R}),\mathrm{d}P_\xi)$ with

$$\begin{aligned} &\int_{\mathbb{R}} J_n^{(\beta-1,\alpha-1)}(s)J_m^{(\beta-1,\alpha-1)}(s)\mathrm{d}P_\xi(s) = \\ &\qquad = \int_{\mathbb{R}} J_n^{(\beta-1,\alpha-1)}(s)J_m^{(\beta-1,\alpha-1)}(s)w(s;\alpha,\beta)\mathrm{d}s \\ &\qquad = \frac{2^{1-\alpha-\beta}}{B(\alpha,\beta)}\int_{-1}^{1} J_n^{(\beta-1,\alpha-1)}(s)J_m^{(\beta-1,\alpha-1)}(s)\tilde{w}(s;\beta-1,\alpha-1)\mathrm{d}s \\ &\qquad = \frac{1}{B(\alpha,\beta)(2n+\alpha+\beta-1)}\frac{\Gamma(n+\beta)\Gamma(n+\alpha)}{\Gamma(n+\alpha+\beta-1)n!}\delta_{nm}. \end{aligned}$$

Moreover, given the nodes s_j and weights $\tilde{\omega}_j$ from the common Gauss-Jacobi-quadrature rule for weighting function $\tilde{w}(.,\beta-1,\alpha-1)$, the Gauss-quadrature rule in terms of weighting function $w(.,\alpha,\beta)$ has the same nodes while the weights are scaled by $\omega_j = \frac{2^{1-\alpha-\beta}}{B(\alpha,\beta)}\tilde{\omega}_j$.

Generalized laguerre polynomials

Generalized Laguerre polynomials are defined by the recurrence relation

$$\begin{aligned}
&La_0^{(\alpha)}(x) = 1,\\
&La_1^{(\alpha)}(x) = 1 + \alpha - x,\\
&(n+1)La_{n+1}^{(\alpha)}(x) = (2n+1+\alpha-x)La_n^{(\alpha)}(x) - (n+\alpha)La_{n-1}^{(\alpha)}(x), n \geq 2
\end{aligned}$$

They form a complete orthogonal system on $L^2([0,\infty), \mathcal{B}([0,\infty)), \tilde{w}(x)\mathrm{d}x)$ with weighting function

$$\tilde{w}(x;\alpha) := x^{\alpha}e^{-x}.$$

More specifically,

$$\int_0^{\infty} La_n^{(\alpha)}(x)La_m^{(\alpha)}(x)\tilde{w}(x;\alpha)\mathrm{d}x = \frac{\Gamma(n+\alpha+1)}{n!}\delta_{nm}.$$

The probability density function of a Gamma-distributed random variable, denoted by $\theta \sim \mathrm{Gamma}(\alpha, \beta)$, with shape parameter α and rate parameter β is given by

$$f_{\theta}(\vartheta;\alpha,\beta) := \frac{\beta^{\alpha}}{\Gamma(\alpha)}\vartheta^{\alpha-1}e^{-\beta\vartheta}\mathbb{1}_{[0,\infty)}(\vartheta).$$

[27] Fixing the transformation between the germ and θ in this case to

$$\psi(s) := \frac{s}{\beta}$$

so that the germ ξ is defined by

$$\xi := \psi^{-1}(\theta) = \beta\theta$$

implies that ξ has probability density function

$$\begin{aligned}
w(s;\alpha,\beta) &= f_{\theta}(\psi(s);\alpha,\beta)\psi'(s) = \frac{\beta^{\alpha}}{\Gamma(\alpha)}\left(\frac{s}{\beta}\right)^{\alpha-1}e^{-s}\frac{1}{\beta}\mathbb{1}_{[0,\infty)}(s)\\
&= \frac{1}{\Gamma(\alpha)}\tilde{w}(s;\alpha-1)\mathbb{1}_{[0,\infty)}(s).
\end{aligned}$$

Since $w(s;\alpha,\beta)$ differs from $\tilde{w}(s;\alpha-1)$ only by a constant factor, it follows that

$$\begin{aligned}
L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{d}P_{\xi}) &= L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), w(s;\alpha,\beta)\mathrm{d}s)\\
&\simeq L^2([0,\infty), \mathcal{B}([0,\infty)), \tilde{w}(s;\alpha-1)\mathrm{d}s),
\end{aligned}$$

[27] We denote by $\Gamma(x)$ the gamma function.

and that the generalized Laguerre polynomials $\{La_n^{(\alpha-1)}\}_{n\in\mathbb{N}_0}$ also form a complete orthogonal system in $L^2(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathrm{d}P_{\tilde{\xi}})$ with

$$\begin{aligned}\int_{\mathbb{R}} La_n^{(\alpha-1)}(s) La_m^{(\alpha-1)}(s)\mathrm{d}P_{\tilde{\xi}}(s) &= \\ &= \int_{\mathbb{R}} La_n^{\alpha-1)}(s) J_m^{(\alpha-1)}(s) w(s;\alpha,\beta)\mathrm{d}s \\ &= \frac{1}{\Gamma(\alpha)}\int_0^{\infty} La_n^{(\alpha-1)}(s) La(\alpha-1)_m(s)\tilde{w}(s;\alpha-1)\mathrm{d}s \\ &= \frac{\Gamma(n+\alpha)}{\Gamma(\alpha)n!}\delta_{nm}.\end{aligned}$$

Moreover, given the nodes s_j and weights $\tilde{\omega}_j$ from the common Gauss-Laguerre-quadrature rule for weighting function $\tilde{w}(.,\alpha-1)$, the Gauss-quadrature rule in terms of weighting function $w(.,\alpha,\beta)$ has the same nodes while the weights are scaled by $\omega_j = \frac{\tilde{\omega}_j}{\Gamma(\alpha)}$.

Appendix 3 Intrusive model expansion

Stochastic Galerkin

For both methods discussed in section 2.1.2, the computation of the expansion coefficients is detached from the underlying procedure from which the model outcome is computed. This is different from the third method. Instead of a more general discussion, we therefore only illustrate this method for the case where the PCE of a model's policy function is constructed. To simplify the notation, suppose that the equations defining the model's solution can be reduced to a sole Euler equation in a single variable. Let $S \subset \mathbb{R}^s$ denote the model's state space and let $g : S \to \mathbb{R}$ denote the variable's policy function. The Euler equation is typically translated into a functional (integral) equation for g, say

$$R(g,x) = 0 \quad \text{for all } x \in S.$$

If the functional equation can not be solved analytically, a common approach is to construct an approximation $\hat{g}$ from linear combinations of some basis functions,[28] say $\Phi_j, j = 1,\ldots,d$, i.e.,

$$\hat{g}(x) = \sum_{j=1}^{d} y_j \Phi_j(x).$$

In order to determine the coefficients y_j in the approximation, which now serves as our model outcome of interest and should not be confused with the Fourier

[28] Most commonly these are selected either as (tensor products of) Chebyshev polynomials or as piecewise linear or cubic polynomials.

coefficients of the PCE, one can, for example, select d appropriate collocation points $x_1, \dots, x_d \in S$ and solve the non-linear system of equations given by

$$R\left(\sum_{j=1}^{d} y_j \Phi_j, x_i\right) = 0 \text{ for all } i = 1, \dots, d$$

for $y_1, \dots, y_d$.

Now consider the case where one parameter is uncertain and hence described by the random variable θ. If the model's (reduced) Euler equation involves θ, then so does the functional equation for g, i.e., we now write

$$R(g, x; \theta) = 0 \quad \text{for all } x \in S.$$

Moreover, if one employs the above-mentioned solution method, the coefficients y_j will typically also depend on θ, i.e., we have, in slight abuse of notation, $Y_j = h_j(\theta)$. In particular, the mappings h_j between the Y_j and θ arise implicitly from the non-linear system of equations

$$R\left(\sum_{j=1}^{d} Y_j \Phi_j, x_i; \theta\right) = 0 \quad \text{for all } i = 1, \dots, d. \tag{13}$$

In order to avoid the necessity for repeated and potentially computationally expensive solutions of this system of equations for different values of θ, one may want to find for each Y_j a PCE in terms of some chosen germ ξ[29]

$$\theta = \psi(\xi) = \sum_{n=0}^{\infty} \hat{\vartheta}_n q_n(\xi),$$

$$Y_j = h_j(\theta) = h_j(\psi(\xi)) = \sum_{n=0}^{\infty} \hat{y}_{jn} q_n(\xi).$$

The PCE of the model's (approximated) policy function with respect to the germ ξ is then given by

$$\hat{g}(x; \xi) = \sum_{j=1}^{d} Y_j \Phi_j(x) = \sum_{j=1}^{d} \sum_{n=0}^{\infty} \hat{y}_{jn} q_n(\xi) \Phi_j(x).$$

Moreover, the Fourier coefficients $\hat{y}_{jn}$ in the PCE can be derived by a Galerkin method if we substitute the Y_j in their implicit definition in (13) with their PCE and impute the corresponding conditions

[29] Note that in this case we have d model outcomes of interest, namely the coefficients $Y_j = h_j(\theta)$ in $\hat{g}$.

$$R\left(\sum_{j=1}^{d}\sum_{n=0}^{\infty}\hat{y}_{jn}q_n(\xi)\Phi_j, x_i;\psi(\xi)\right) = 0 \text{ in } L^2, \forall i = 1,\dots,d$$

$$\Leftrightarrow \left\langle R\left(\sum_{j=1}^{d}\sum_{n=0}^{\infty}\hat{y}_{jn}q_n(\xi)\Phi_j, x_i;\psi(\xi)\right), q_m(\xi)\right\rangle_{L^2} = 0 \ \forall i = 1,\dots,d, \ \forall m \in \mathbb{N}_0.$$

Hence, we can solve for the $d(N+1)$ unknown coefficients $\hat{y}_{jn}$ in the truncated PCE $Y_j \approx \sum_{n=0}^{N}\hat{y}_{jn}q_n(\xi)$ from the system of equations

$$0 \approx \left\langle R\left(\sum_{j=1}^{d}\sum_{n=0}^{N}\hat{y}_{jn}q_n(\xi)\Phi_j, x_i;\psi(\xi)\right), q_m(\theta)\right\rangle_{L^2}$$
$$= \int_{\mathbb{R}} R\left(\sum_{j=1}^{d}\sum_{n=0}^{N}\hat{y}_{jn}q_n(\xi)\Phi_j, x_i;\psi(\xi)\right) q_m(\xi)\mathrm{d}P_{\xi}(\xi)$$

for $i = 1,\dots,d$ and $m = 0,\dots,N$. The integral is computed numerically, either from Monte-Carlo draws or from an appropriate Gauss quadrature. Moreover, $\psi(\xi)$ can be substituted by its truncated series expansion as previously described in subsection 2.1.1.

Appendix 4 Smolyak-Gauss-Quadrature

Suppose that for every $i = 1,\dots,k$ the distribution P_{ξ_i} of ξ_i possesses a probability density function w_i, so that $w := \prod_{i=1}^{k} w_i$ is the probability density of $P_{\boldsymbol{\xi}}$. Then (6b) becomes

$$\hat{y}_{\alpha} = \|q_{\alpha}\|_{L^2}^{-2} \quad \times$$
$$\int_{\mathbb{R}}\dots\int_{\mathbb{R}} h(\psi(s_1,\dots,s_k))q_{1\alpha_1}(s_1)\dots q_{k\alpha_k}(s_k)w_1(s_1)\dots w_k(s_k)\mathrm{d}s_1\dots\mathrm{d}s_k. \tag{14}$$

Further, suppose that one-dimensional Gauss-quadrature rules corresponding to weighting functions w_i and orthogonal polynomials $\{q_{in}\}_{n\in\mathbb{N}_0}$ are available. For $i = 1,\dots,k$ let $Q_i(M_i)$ denote this one-dimensional Gauss-quadrature rule with M_i nodes $\{s_{i,M_i}^{(j)}\}_{j=1,\dots,M_i}$ and weights $\{\omega_{i,M_i}^{(j)}\}_{j=1,\dots,M_i}$, i.e.,

$$Q_i(M_i)g := \sum_{j=1}^{M_i}\omega_{i,M_i}^{(j)}g(s_{i,M_i}^{(j)}) \ \text{ for } g \in L_i^2.$$

Then choose for each $i = 1,\dots,k$ an increasing sequence of natural numbers $\{M_{ij}\}_{j\in\mathbb{N}} \subset \mathbb{N}, M_{ij+1} > M_{ij}$ and define the difference operator by

$$\Delta_{i1} := Q_i(M_{i1}) \quad \text{and} \quad \Delta_{ij} := Q_i(M_{ij}) - Q_i(M_{ij-1}),\ j \geq 2.$$

The Smolyak-Gauss-quadrature rule of order $l \in \mathbb{N}$ and with growth rules given by $\{M_{ij}\}_{j\in\mathbb{N}}$ is defined by

$$Q_l := \sum_{\substack{\nu \in \mathbb{N}^k \\ |\nu| \leq k+l}} \bigotimes_{i=1}^{k} \Delta_{iv_i}.$$

or equivalently taking care of duplicate terms in the difference operators

$$Q_l = \sum_{\substack{\nu \in \mathbb{N}^k \\ \max\{k,l+1\} \leq |\nu| \leq k+l}} (-1)^{k+l-1} \binom{k-1}{k+l-|\nu|} \bigotimes_{i=1}^{k} Q_i(M_{iv_i}).$$

Applying the Smolyak-Gauss-quadrature rule to (14) in particular yields the approximation

$$\begin{aligned}\hat{y}_\alpha \approx \left(\prod_{i=1}^{k} \|q_{i\alpha_i}\|^2_{L_i^2}\right)^{-1} \sum_{\substack{\nu \in \mathbb{N}^k \\ \max\{k,l+1\} \leq |\nu| \leq k+l}} (-1)^{k+l-1} \binom{k-1}{k+l-|\nu|} \\ \sum_{j_1=1}^{M_{1,\nu_1}} \cdots \sum_{j_k=1}^{M_{k,\nu_k}} \omega^{(j_1)}_{1,M_{1,\nu_1}} \cdots \omega^{(j_k)}_{k,M_{k,\nu_k}} \\ \times h\left(\psi\left(s^{(j_1)}_{1,M_{1,\nu_1}} \ldots s^{(j_k)}_{k,M_{k,\nu_k}}\right)\right) \\ \times q_{1\alpha_1}\left(s^{(j_1)}_{1,M_{1,\nu_1}}\right) \cdots q_{k\alpha_k}\left(s^{(j_k)}_{k,M_{k,\nu_k}}\right).\end{aligned}$$

This procedure requires to evaluate the model outcome of interest $h\left(\psi\left(s^{(j_1)}_{1,M_{1,\nu_1}} \ldots s^{(j_k)}_{k,M_{k,\nu_k}}\right)\right)$ at all sparse-grid points.

Appendix 5 Monomial rules

Stroud (1971) introduces sparse numerical integration with monomial rules and presents various rules to integrate in different spaces. In this section, we present some numerical results for the calculation of the PCE coefficients.

Rosenbrock function

To show the general functioning of the monomial quadrature rules, we first replicate the exercise of Bhusal and Subbarao (2020), i.e., approximate the Rosenbrock function

Table 11 Rosenbrock PCE approximation

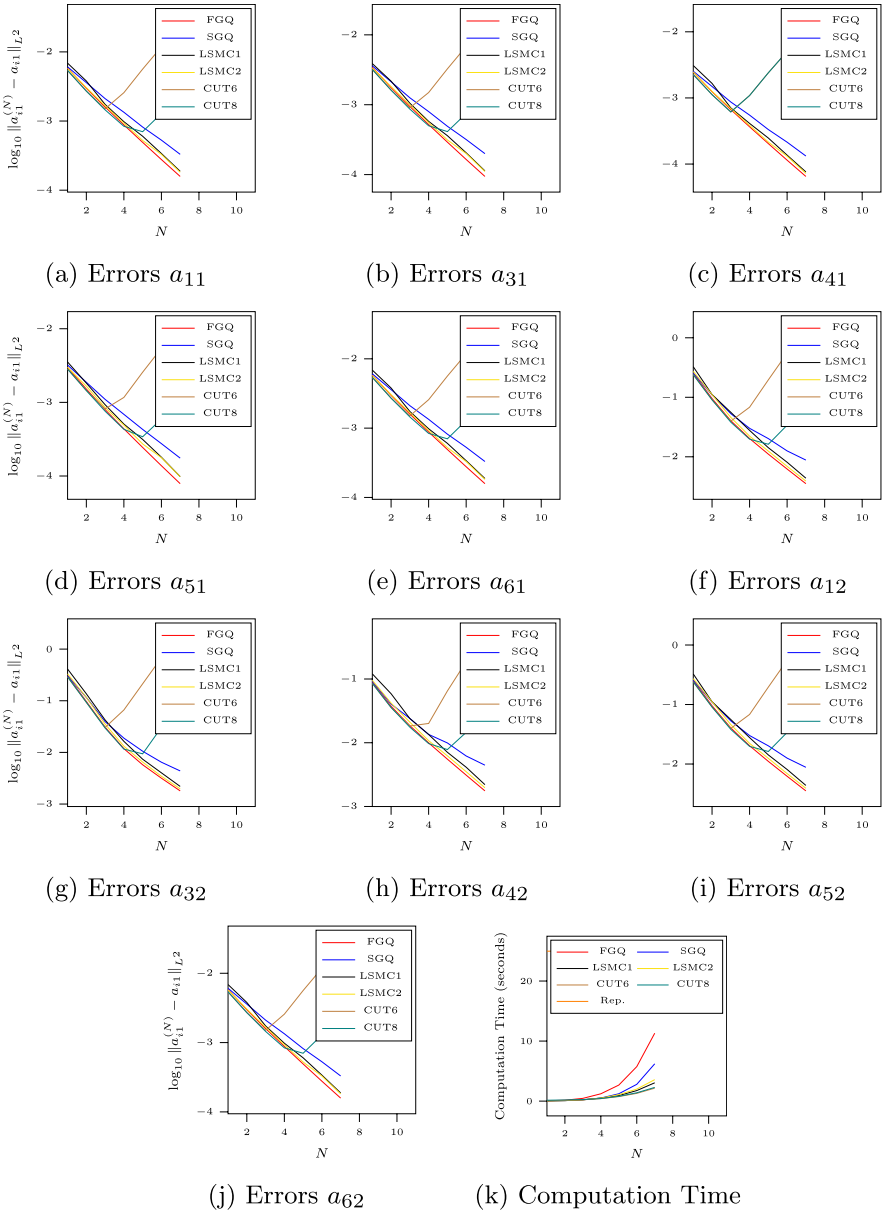
5-d Rosenbrock PCEapproximation						
Trunc. lvl	4		5		6	
	$\log_{10} L^2$ Error	N_{Grid}	$\log_{10} L^2$ Error	N_{Grid}	$\log_{10} L^2$ Error	N_{Grid}
Tensor Grid	−10.98	3,125	−10.87	7,776	−11.15	16,807
Sparse Grid	−9.74	781	−9.70	781	−9.37	2,203
Least Squares	−10.91	252	−10.81	504	−10.58	924
CUT-6	1.90	155	2.49	155	3.20	155
CUT-8	−10.44	425	−10.45	425	1.45	425
6-d Rosenbrock PCE approximation						
Trunc. lvl	4		5		6	
	$\log_{10} L^2$ Error	N_{Grid}	$\log_{10} L^2$ Error	N_{Grid}	$\log_{10} L^2$ Error	N_{Grid}
Tensor Grid	−10.87	15,625	−10.72	46,656	−10.84	117,649
Sparse Grid	−9.31	1,433	−9.28	1,433	−8.71	4,541
Least Squares	−10.94	420	−10.77	924	−10.60	1,848
CUT-6	2.07	301	2.61	301	3.35	301
CUT-8	−6.93	973	−6.89	973	1.79	937

Tensor grid lvl. = Trunc. lvl.=4 +1, Smolyak, min N_{Grid}, given $\log_{10} L^2$ Error<-5, Least Squares, twice PCE coefficients

$$f(x) = \sum_{i=1}^{d-1} 100\left(x_{i+1} - x_i^2\right)^2 + (1 - x_i)^2$$

with PCE. We consider the cases where $x_i \sim U(-2, 2)$ and $d \in \{5, 6\}$. As Bhusal and Subbarao (2020), we consider the CUT-8 and CUT-6 rules from Adurthi et al. (2018) and full tensor grid, sparse grid, and least squares from the main body of the paper. We consider a truncation at levels 4, 5, and 6. Lastly, note that for those dimensions ($d \in \{5, 6\}$), we could not find any monomial rules presented by Stroud (1971) higher degree 5 that have solely non-negative weights and are in the variables space, e.g., the first weight of the fifth-degree rule presented in Judd (1998) (Stroud (1971) $C_n 5 - 5$) becomes −60.44 for $d = 5$. Further, the degree 5 rules with solely positive weights approximate the Rosenbrock function poorly. Lastly, the CUT-8 rule nodes leave the boundaries of space of x_i for $d > 6$ and has already one negative weight for $d = 6$.

Table 11 presents the results. The CUT-8 rule performs well for $d = 5$. However, the CUT-8 rule is outperformed by Least Squares. The performance of the CUT-8 becomes worse with $d = 6$, where one weight becomes negative ($\approx -.5$), yet the approximation seems still good.



(a) Errors a_{11} (b) Errors a_{31} (c) Errors a_{41}

(d) Errors a_{51} (e) Errors a_{61} (f) Errors a_{12}

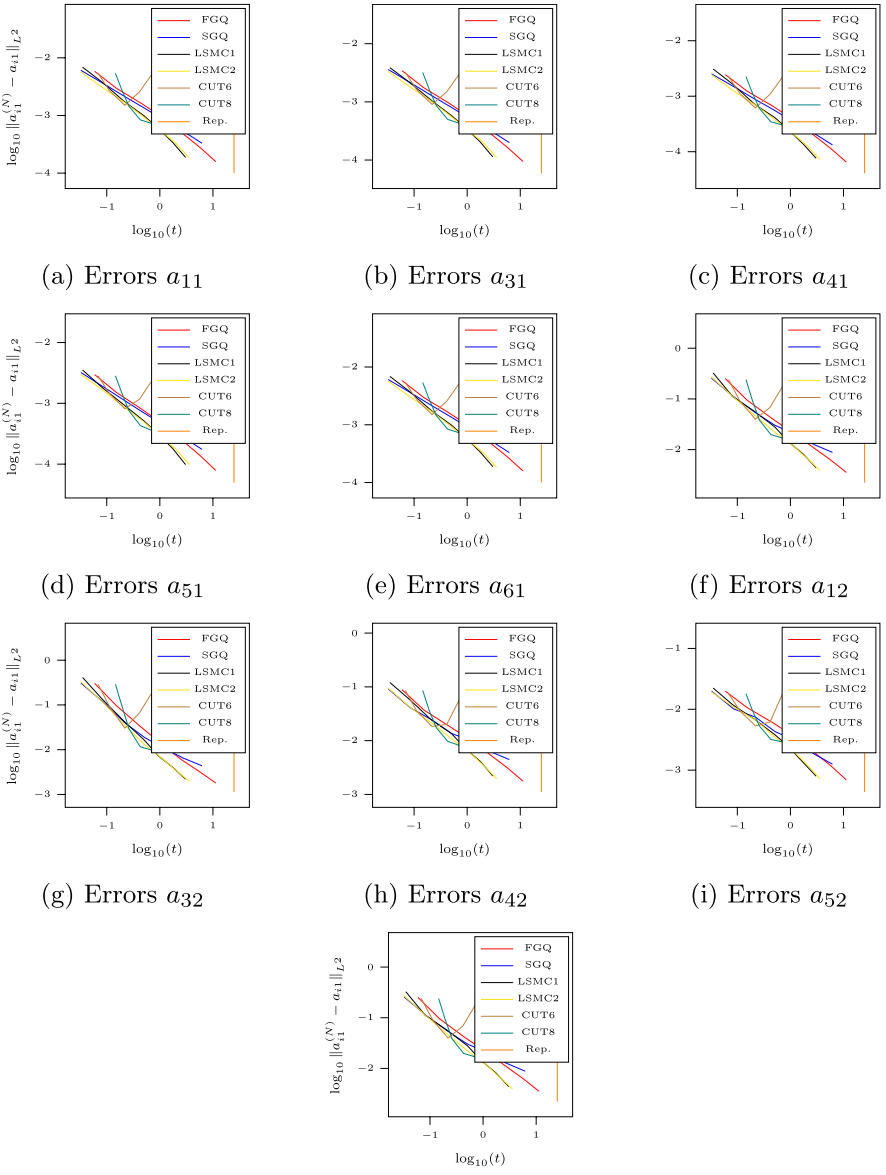
(g) Errors a_{32} (h) Errors a_{42} (i) Errors a_{52}

(j) Errors a_{62} (k) Computation Time

Fig. 10 L^2 Convergence of PCE with approximated coefficients and computation time on an Intel® Core™i7-7700 CPU @ 3.60GHz

RBC Model

Now we replicate the integration analysis of the main paper (Figure 4 and 5 there) for the CUT rules. Given the results of the previous section on monomial rules, we



(a) Errors a_{11} (b) Errors a_{31} (c) Errors a_{41}

(d) Errors a_{51} (e) Errors a_{61} (f) Errors a_{12}

(g) Errors a_{32} (h) Errors a_{42} (i) Errors a_{52}

(j) Errors a_{62}

Fig. 11 L^2 Convergence of PCE with approximated coefficients and computation time on an Intel® Core™i7-7700 CPU @ 3.60GHz

reduce the space to 5 dimensions (β is now fixed) and assume $\frac{\theta-\hat{\varphi}_0}{\hat{\varphi}_1} = \psi(s) \sim B(1,1) = U(0,1)$ for all θ.

Figures 10 and 11 illustrate the analyses. It turns out that the monomial rule CUT8 outperforms all other sparse methods for truncation at $N = 5$ and all methods in time at this truncation level. However, a higher truncation leads to more imprecise approximations. The problem is the lack of high-degree, high-dimensional monomial rules for different distributions. However, suitable cases for existing monomial rules seem to work well, which motivates further research to find high-degree, high-dimensional monomial rules for mixed distributions.

Appendix 6 Further applications of Generalized Polynomial Chaos Expansions

We present here additional applications of PCE. First, we show how to use PCE as surrogates for the gradients. Further, statistical properties of the model outcome, as induced by the predefined distribution of the uncertain input parameters, can be derived directly from the PCE. Additionally, in somewhat other contexts, PCE can be used to discretize the space of cross-sectional distributions.

Surrogate for Gradients

The truncated PCE in (9) may also be used to approximate the derivatives of the mapping h between parameter values and model outcomes. More specifically, the PCE provides the approximation

$$\frac{\partial h}{\partial \vartheta_i}(\vartheta) \approx \sum_{\alpha \in \mathbb{N}_0^k, |\alpha| \leq N} \hat{y}_\alpha \sum_{j=1}^{k} \frac{\partial q_\alpha}{\partial s_k}(\psi^{-1}(\vartheta)) \frac{\partial \psi_j^{-1}}{\partial \vartheta_i}(\vartheta).$$

This approximation can be useful if such derivatives must be evaluated at a potentially large number of points. One example may be the method proposed by Iskrev (2010) for conducting local identification analysis which requires differentiation of the linearized policy function concerning the parameters.

Evaluation of Statistical Properties

Convergence in $L^2(\Omega, \mathcal{A}, P)$ of the series expansion in (5b) implies that the distribution of the model outcome Y can be equivalently characterized by its polynomial expansion. In particular, the mean and variance of Y follow directly from the fact that convergence in L^2 also implies convergence of the mean and variance so that orthogonality of the polynomials (and $q_{\mathbf{0}} = 1$ for $\mathbf{0} := (0, \ldots, 0) \in \mathbb{N}_0^k$) yields

$$\mathbb{E}[Y] = \sum_{\alpha \in \mathbb{N}_0^k} \hat{y}_\alpha \mathbb{E}[q_\alpha(\boldsymbol{\xi})] = \sum_{\alpha \in \mathbb{N}_0^k} \hat{y}_\alpha \mathbb{E}[q_\alpha(\boldsymbol{\xi}) q_{\mathbf{0}}(\boldsymbol{\xi})] = \sum_{\alpha \in \mathbb{N}_0^k} \hat{y}_\alpha \langle q_\alpha, q_{\mathbf{0}} \rangle_{L^2} = \hat{y}_{\mathbf{0}},$$

and

$$\mathrm{Var}[Y] = \mathbb{E}\left[\left(\sum_{\alpha\in\mathbb{N}_0^k} \hat{y}_\alpha q_\alpha(\boldsymbol{\xi}) - \hat{y}_{\mathbf{0}}\right)^2\right] = \mathbb{E}\left[\left(\sum_{\alpha\in\mathbb{N}_0^k\setminus\{\mathbf{0}\}} \hat{y}_\alpha q_\alpha(\boldsymbol{\xi})\right)^2\right] =$$
$$= \sum_{\alpha,\beta\in\mathbb{N}_0^k\setminus\{\mathbf{0}\}} \hat{y}_\alpha\hat{y}_\beta\langle q_\alpha, q_\beta\rangle_{L^2} = \sum_{\alpha\in\mathbb{N}_0^k\setminus\{\mathbf{0}\}} \hat{y}_\alpha^2\|q_\alpha\|_{L^2}^2.$$

Moreover, other statistical properties can be computed by Monte Carlo methods. Large samples of Y can be efficiently constructed by drawing from the germ's distribution and inserting the sample into the expansion of Y. Compared to traditional methods, repeated and costly model evaluations can thus be avoided.

Sobol's indices for global variance-based sensitivity analysis The decomposition of the model's outcome variance from above also lays the foundation for the sensitivity analyses of Harenberg et al. (2019). More specifically, consider a truncated PCE $S_N^{\mathrm{tot}}(Y)$ or $S_N^{\max}(Y)$ of the model outcome Y as in (7b) or (8b). By reordering, one can then equivalently write the truncated PCE as

$$S_N^{\mathrm{tot}}(Y) = \sum_{I\subset\{1,\dots,k\}} \sum_{\substack{\alpha\in\mathbb{N}_0^k,\,|\alpha|\le N\\ \alpha_i\neq 0\ \forall i\in I\\ \alpha_i = 0\ \forall i\notin I}} \hat{y}_\alpha q_\alpha(\xi),$$

i.e., for any collection $\{\xi_i\}_{i\in I}$ where $I\subset\{1,\dots,k\}$ we now explicitly group the polynomials $q_\alpha(\boldsymbol{\xi})$ with non-zero degree in each $\xi_i, i\in I$ but zero-degree in all $\xi, i\notin I$. Orthogonality of the polynomials then implies for any nonempty collection $I\subset\{1,\dots,k\}, I\neq\emptyset$ that

$$V_I := \mathrm{Var}\Big[\sum_{\substack{\alpha\in\mathbb{N}_0^k,\,|\alpha|\le N\\ \alpha_i\neq 0\ \forall i\in I\\ \alpha_i = 0\ \forall i\notin I}} \hat{y}_\alpha q_\alpha(\boldsymbol{\xi})\Big] = \sum_{\substack{\alpha\in\mathbb{N}_0^k,\,|\alpha|\le N\\ \alpha_i\neq 0\ \forall i\in I\\ \alpha_i = 0\ \forall i\notin I}} \hat{y}_\alpha^2\|q_\alpha\|_{L^2}^2$$

and

$$V := \mathrm{Var}[S_N^{\mathrm{tot}}(Y)] = \sum_{\substack{I\subset\{1,\dots,k\},\\ I\neq\emptyset}} V_I.$$

The Sobol indices then describe the shares of the variance that are explained by a collection $\{\xi_i\}_{i\in I}$ of germs for $I\subset\{1,\dots,k\}, I\neq\emptyset$

$$S_I := \frac{V_I}{V}.$$

The first order Sobol indices $S_{\{i\}}$ for single germs ξ_i are interpreted as the fraction of the total variance which would disappear when ξ_i would be perfectly known. On the other hand, the total contribution indices are defined by

$$S_i^T := \frac{\sum^{I \subset \{1,\dots,k\},}_{i \in I} V_I}{V}$$

and describe the germ's total contribution to the outcome's variance.

Relevance for (Bayesian) estimation As Harenberg et al. (2019) note, a sufficient size of the total Sobol' index of the parameter ϑ_i is a necessary condition for the identifiability of ϑ_i using Y. In terms of Bayesian estimation, PCE also facilitates the comparison of the model outcome's prior and posterior distribution. Once we have obtained the parameters' posterior distribution, PCE enables the representation of the corresponding posterior distribution of the model's outcome. We can then compare the PCE implied variances and the contribution of an arbitrary set of parameters, which delivers an indicator for the reduced uncertainty of the model outcome Y subject to this set of parameters.

Discretizing space of cross-sectional distributions

Here, we briefly present the possibility of using PCE to discretize the state space in models where a cross-sectional distribution over heterogeneous agents becomes a state variable for individual decision rules as suggested by Pröhl (2017). Her examples are models that combine idiosyncratic income risk with aggregate productivity risk as Aiyagari (1994). In such models, households need to know the decision rules of other households to form rational expectations about future aggregates and prices for their own decisions. Yet, since the decisions of other households depend on their respective individual states, households need to factor in the whole cross-sectional distribution over individual states for their own decision. In consequence, the cross-sectional distribution of individual states becomes an argument for the individuals' policy function in such models.

The literature offers different approaches in order to discretize the state space. Krusell and Smith (1998) suggest a bounded rationality approach and base the individuals' policy function only on partial information from the cross-sectional distribution, e.g., a finite number of moments, and a parametric law of motion for these measures. The method of Reiter (2009) discretizes the state space by piecewise uniform distributions over a finite number of histogram bins. Differently, Pröhl (2017) replaces the cross-sectional distribution as an argument of the decision rule by the coefficients of its (truncated) PCE given a choice of germs. More precisely, if ξ denotes the germ with cumulative distribution function F_ξ and μ_t is the cross-sectional distribution over individual states in period t, then the random variable

$$\theta_t := \mu_t^{-1} \circ F_\xi \circ \xi$$

is distributed according to μ_t.[30] One can then compute the coefficients $\hat{\vartheta}_{n,t}$ of its PCE

[30] Note that θ_t does not denote a model parameter in this context as in the rest of the present paper. Instead, θ_t is a random variable that is distributed according to the cross-sectional distribution μ_t and that is a function of the germ. Hence, θ_t can be interpreted as the random variable constructed from the basis ξ that describes a random draw from the mass of heterogenous agents in period t.

$$\theta_t = \mu_t^{-1}(F_\xi(\xi)) = \sum_{n=0}^{\infty} \hat{\vartheta}_{n,t} q_n(\xi)$$

analog to the methods described in section 2 from

$$\hat{\vartheta}_{n,t} = \|q_n\|_{L^2}^{-2} \langle \mu_t^{-1} \circ F_\xi, q_n \rangle_{L^2} = \|q_n\|_{L^2}^{-2} \int_{\mathbb{R}} (\mu_t^{-1} \circ F_\xi) q_n \mathrm{d}P_\xi.$$

Instead of the cross-sectional distribution μ_t, one can then use a finite number of the PCE coefficients $\hat{\vartheta}_{n,t}$ as arguments of the individual policy function. On the one hand, the PCE coefficients $\hat{\vartheta}_{n,t}$ then allow to recover the cross-sectional distribution μ_t and aggregate variables in period t. On the other hand, the individual decision rules imply the law of motion of the cross-sectional distribution, $\mu_t \mapsto \mu_{t+1}$, and the PCE coefficients $\hat{\vartheta}_{n,t+1}$ of μ_{t+1} can be derived as above. Hence, the method of Pröhl (2017) does not need a parametric assumption about the law of motion for the cross-sectional distribution. Pröhl (2017) shows that this approach provides more precise solutions and thereby, brings new economic characteristics of those well-known models.

Appendix 7 Supplementary Results

Figure 12 contain the supplement MLE results for a sample size of $T = 500$ and a particle filter with 10000 particles, respectively. Analogously, to Tables 7 and 8 the Tables 12, 13 and 14 in this appendix contains the full BE results based on the global projection solution.

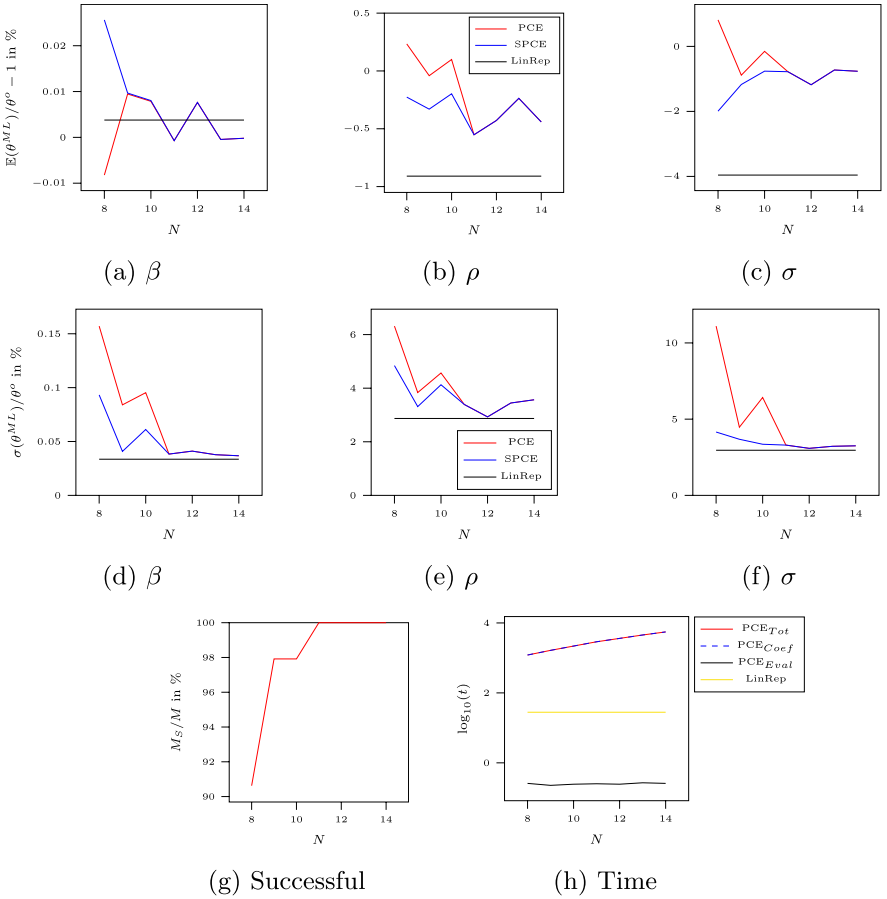


Fig. 12 ML from various likelihood approximations ($M = 96$) with T=500. PCE: PCE approximated likelihood from a particle filter, SPCE: Only successful PCE approximations (M_S), i.e., exclusion of maxima at the parameter bounds. LinRep: Repeated likelihood evaluation using the Kalman Filter from the linear model solution. N equals the truncation level, the quadrature level equals N+1. Computation time on one core of an AMD® EPYC™7313 (Milan) CPU @ 3.00GHz

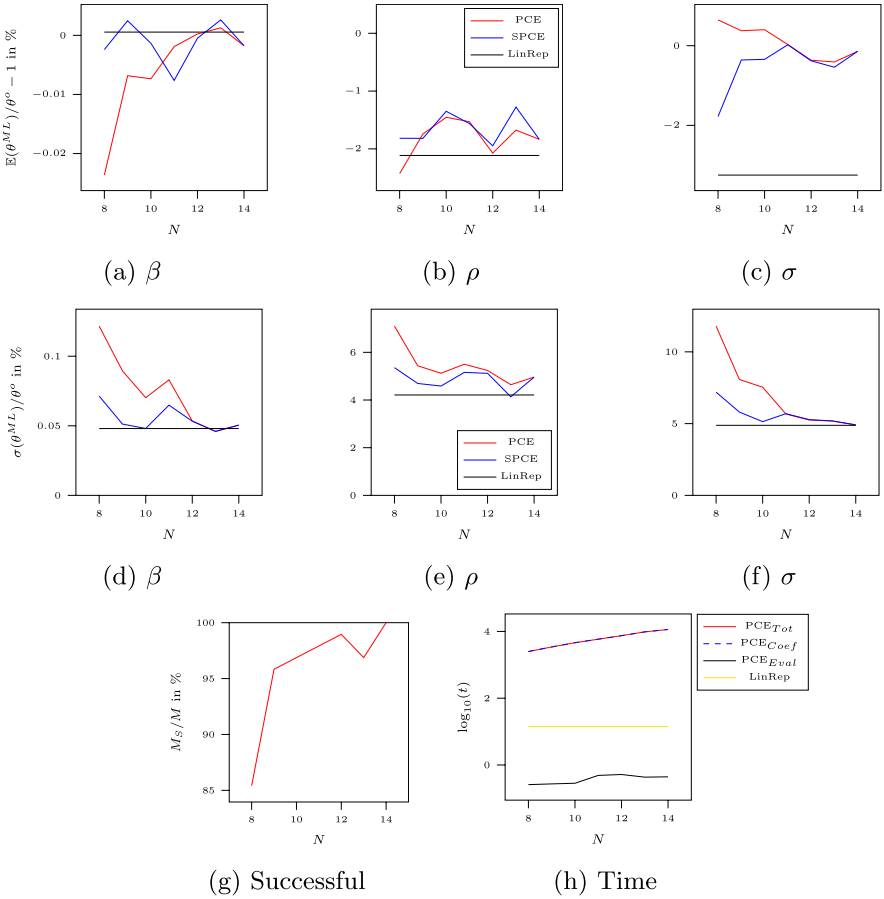


Fig. 13 ML from various likelihood approximations ($M = 96$) with 10,000 particles. PCE: PCE approximated likelihood from a particle filter, SPCE: Only successful PCE approximations (M_S), i.e., exclusion of maxima at the parameter bounds. LinRep: Repeated likelihood evaluation using the Kalman Filter from the linear model solution. N equals the truncation level, the quadrature level equals N+1. Computation time on one core of an AMD® EPYC™7313 (Milan) CPU @ 3.00GHz

Table 12 Monte Carlo results - Bayesian estimation (global) I

Benchmark (repeated solution)									
Time:	Total average 18:17:37								
Linear policy function									
Time:	Total average 00:05:45								
		Mean:				Quantile:			
j	x:		5%	10%	25%	50%	75%	90%	95%
β	$\bar{\epsilon}_j(x)$	1.70	2.30	2.01	1.75	1.62	1.70	1.92	2.17
	$\epsilon_j(x)_{.05}$	0.14	0.08	0.09	0.15	0.10	0.15	0.09	0.10
	$\epsilon_j(x)_{.5}$	1.46	1.77	1.69	1.66	1.40	1.46	1.55	1.84
	$\epsilon_j(x)_{.95}$	3.63	5.87	4.98	4.06	3.44	4.12	5.03	5.83
ρ	$\bar{\epsilon}_j(x)$	3.68	3.57	3.66	3.80	3.87	3.86	3.78	3.71
	$\epsilon_j(x)_{.05}$	0.19	0.13	0.07	0.32	0.25	0.19	0.28	0.29
	$\epsilon_j(x)_{.5}$	3.64	2.96	3.21	3.62	3.77	3.53	3.25	3.29
	$\epsilon_j(x)_{.95}$	9.07	8.81	8.80	9.14	9.61	8.82	8.56	8.35
σ	$\bar{\epsilon}_j(x)$	2.56	3.51	3.35	3.08	2.70	2.21	1.80	1.63
	$\epsilon_j(x)_{.05}$	0.72	2.03	1.90	1.58	0.87	0.37	0.24	0.21
	$\epsilon_j(x)_{.5}$	2.49	3.47	3.31	3.05	2.69	2.12	1.67	1.57
	$\epsilon_j(x)_{.95}$	4.57	5.18	5.05	4.88	4.66	4.40	4.12	3.75

Observable: Output Y_t. $\bar{\epsilon}_j$: mean error, $\epsilon_{j,.05}$: 5 percentile of error, $\epsilon_{j,.5}$: median of error, $\epsilon_{j,.95}$: 95 percentile of error. Errors are expressed as deviations from the benchmark method of repeatedly solving the (global) policy function in percent of the range of the parameter's distribution. Time: hh:mm:ss on one core of an AMD® EPYC™7313 (Milan) CPU @ 3.00GHz

Table 13 Monte Carlo results - Bayesian estimation (global) II

PCE Posterior-Kernel									
Time:		Total average 00:32:27			PCE 00:32:24		Estimation average 00:00.03		
		Mean:				Quantile:			
j	x:		5%	10%	25%	50%	75%	90%	95%
β	$\bar{\epsilon}_j(x)$	1.63	2.04	1.98	1.86	1.75	1.68	1.64	1.71
	$\epsilon_j(x)_{.05}$	0.09	0.09	0.06	0.19	0.06	0.06	0.15	0.12
	$\epsilon_j(x)_{.5}$	0.80	0.90	1.01	0.81	0.99	0.90	0.89	0.86
	$\epsilon_j(x)_{.95}$	3.08	4.31	4.10	3.95	3.51	3.38	3.53	3.64
ρ	$\bar{\epsilon}_j(x)$	2.56	2.85	3.07	3.03	3.09	2.71	2.70	2.58
	$\epsilon_j(x)_{.05}$	0.11	0.28	0.22	0.25	0.16	0.10	0.14	0.09
	$\epsilon_j(x)_{.5}$	1.75	1.96	2.41	1.94	2.31	1.70	1.99	1.88
	$\epsilon_j(x)_{.95}$	7.17	6.79	7.90	8.55	8.46	7.65	7.44	6.75
σ	$\bar{\epsilon}_j(x)$	0.70	0.77	0.76	0.75	0.77	0.73	0.75	0.78
	$\epsilon_j(x)_{.05}$	0.02	0.05	0.08	0.06	0.04	0.03	0.07	0.06
	$\epsilon_j(x)_{.5}$	0.51	0.56	0.50	0.54	0.57	0.49	0.51	0.49
	$\epsilon_j(x)_{.95}$	2.10	1.82	1.98	2.07	2.21	2.30	2.64	2.38

Observable: Output Y_t. $\bar{\epsilon}_j$: mean error, $\epsilon_{j,.05}$: 5 percentile of error, $\epsilon_{j,.5}$: median of error, $\epsilon_{j,.95}$: 95 percentile of error. Errors are expressed as deviations from the benchmark method of repeatedly solving the (global) policy function in percent of the range of the parameter's distribution. Time: hh:mm:ss on one core of an AMD® EPYC™7313 (Milan) CPU @ 3.00GHz. The truncation degree and quadratur level of the expanded QoI are 14 and 13, respectively

Table 14 Monte Carlo results - Bayesian estimation (global) III

PCE policy function(global)									
Time:		Total average 17:45:11			PCE 00:00.27		Estimation average 17:44:44		
		Mean:			Quantile:				
j	x:		5%	10%	25%	50%	75%	90%	95%
β	$\bar{\epsilon}_j(x)$	0.09	0.25	0.15	0.09	0.06	0.07	0.13	0.21
	$\epsilon_j(x)_{.05}$	0.00	0.02	0.02	0.01	0.00	0.00	0.01	0.01
	$\epsilon_j(x)_{.5}$	0.05	0.18	0.12	0.07	0.05	0.06	0.11	0.16
	$\epsilon_j(x)_{.95}$	0.26	0.71	0.35	0.22	0.14	0.20	0.32	0.67
ρ	$\bar{\epsilon}_j(x)$	0.20	0.32	0.28	0.24	0.23	0.27	0.31	0.43
	$\epsilon_j(x)_{.05}$	0.01	0.03	0.02	0.01	0.01	0.04	0.00	0.03
	$\epsilon_j(x)_{.5}$	0.17	0.24	0.23	0.22	0.18	0.23	0.28	0.37
	$\epsilon_j(x)_{.95}$	0.48	0.95	0.64	0.57	0.58	0.66	0.78	1.08
σ	$\bar{\epsilon}_j(x)$	0.05	0.11	0.09	0.08	0.07	0.08	0.11	0.15
	$\epsilon_j(x)_{.05}$	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.03
	$\epsilon_j(x)_{.5}$	0.04	0.10	0.07	0.07	0.06	0.06	0.10	0.13
	$\epsilon_j(x)_{.95}$	0.14	0.25	0.21	0.16	0.18	0.20	0.27	0.35

Observable: Output Y_t. $\bar{\epsilon}_j$: mean error, $\epsilon_{j,.05}$: 5 percentile of error, $\epsilon_{j,.5}$: median of error, $\epsilon_{j,.95}$: 95 percentile of error. Errors are expressed as deviations from the benchmark method of repeatedly solving the (global) policy function in percent of the range of the parameter's distribution. Time: hh:mm:ss on one core of an AMD® EPYC™7313 (Milan) CPU @ 3.00GHz. The truncation degree and quadratur level of the expanded QoI are 14 and 13, respectively

Acknowledgements We are grateful to our academic advisor Prof. Alfred Maußner for his various insightful and detailed remarks on different versions of this article. We thank Michel Bauer for his initial ideas and continuing discussions. This research was supported in part through high-performance computing resources available at the Kiel University Computing Centre which is partly founded by the German Research Foundation (DFG) - project number 440395346. This research was also supported by the DFG through grant nr. 465135565 "Models of Imperfect Rationality and Redistribution in the context of Retirement"

Funding Open Access funding enabled and organized by Projekt DEAL. The authors have not disclosed any funding.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Adurthi, N., Singla, P., & Singh, T. (2018). Conjugate unscented transformation: Applications to estimation and control. Journal of Dynamic Systems, Measurement, and Control 140(3):030,907

Aiyagari, S. R. (1994). Uninsured idiosyncratic risk and aggregate saving. *The Quarterly Journal of Economics, 109*(3), 659–684.

Albeverio, S., Cordoni, F., Di Persio, L., et al. (2019). Asymptotic expansion for some local volatility models arising in finance. *Decisions in Economics and Finance, 42*, 527–573.

Bhusal, R., & Subbarao, K. (2020). Generalized polynomial chaos expansion approach for uncertainty quantification in small satellite orbital debris problems. *The Journal of the Astronautical Sciences, 67*, 225–253.

Blanchard, O. J., & Kahn, C. M. (1980). The Solution of linear difference models under rational expectations. *Econometrica, 48*(5), 1305–1311.

Bürkner, P. C., Kröker, I., Oladyshkin, S., et al. (2023). A fully bayesian sparse polynomial chaos expansion approach with joint priors on the coefficients and global selection of terms. *Journal of Computational Physics, 488*(112), 210.

Cameron, R. H., & Martin, W. T. (1947). The orthogonal development of non-linear functionals in series of Fourier-Hermite functionals. *Annals of Mathematics, 48*(2), 385–392.

Cheng, K., & Lu, Z. (2018). Adaptive sparse polynomial chaos expansions for global sensitivity analysis based on support vector regression. *Computers and Structures, 194*, 86–96.

Dias, F. S., & Peters, G. W. (2021). Option pricing with polynomial chaos expansion stochastic bridge interpolators and signed path dependence. *Applied Mathematics and Computation, 411*(126), 484.

Fernández-Villaverde, J., & Rubio-Ramírez, J. F. (2005). Estimating dynamic equilibrium economies: linear versus nonlinear likelihood. *Journal of Applied Econometrics, 20*(7), 891–910.

Gersbach, H., Liu, Y., & Tischhauser, M. (2021). Versatile forward guidance: Escaping or switching? *Journal of Economic Dynamics and Control, 127*(104), 087.

Ghanem, R. G., & Spanos, P. D. (1991). Spectral stochastic finite-element formulation for reliability analysis. *Journal of Engineering Mechanics, 117*(10), 2351–2372.

Harenberg, D., Marelli, S., Sudret, B., et al. (2019). Uncertainty quantification and global sensitivity analysis for economic models. *Quantitative Economics, 10*(1), 1–41.

Heer, B., & Maußner, A. (2024). *Weighted residuals methods*. Cham: Springer.

Herbst, E. P., & Schorfheide, F. (2016). *Bayesian estimation of DSGE models. Oxford: The Econometric and Tinbergen Institutes lectures*. Princeton: Princeton University Press.

Iskrev, N. (2010). Local identification in DSGE models. *Journal of Monetary Economics, 57*(2), 189–202.

Jackson, D. (1941). Fourier series and orthogonal polynomials. Mathematical Association of America

Jacquelin, E., Adhikari, S., Sinou, J. J., et al. (2015). Polynomial chaos expansion in structural dynamics: Accelerating the convergence of the first two statistical moment sequences. *Journal of Sound and Vibration, 356*, 144–154.

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

Judd, KL. (1996). Approximation, perturbation, and projection methods in economic analysis. In: Amman HM, Kendrick DA, Rust J (eds) Handbook of Computational Economics, Handbook of Computational Economics, vol 1. Elsevier, chap 12, p 509–585

Judd, K. L. (1998). *Numerical methods in economics*. Cambridge: MIT Press.

Kaintura, A., Dhaene, T., & Spina, D. (2018). Review of polynomial chaos-based methods for uncertainty quantification in modern integrated circuits. *Electronics, 7*(3), 30. https://doi.org/10.3390/electronics7030030

Klein, P. (2000). Using the generalized Schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control, 24*(10), 1405–1423.

Krusell, P., & Smith, A. A. J. (1998). Income and wealth heterogeneity in the macroeconomy. *Journal of Political Economy, 106*(5), 867–896.

Lu, F., Morzfeld, M., Tu, X., et al. (2015). Limitations of polynomial chaos expansions in the bayesian solution of inverse problems. *Journal of Computational Physics, 282*, 138–147. https://doi.org/10.1016/j.jcp.2014.11.010

Lüthen, N., Marelli, S., & Sudret, B. (2021). Sparse polynomial chaos expansions: Literature survey and benchmark. *SIAM/ASA Journal on Uncertainty Quantification, 9*(2), 593–649. https://doi.org/10.1137/20M1315774

Marconi, D. (2016). Polynomial chaos expansion approach to malliavin calculus analysis of bond option sensitivity. *International Journal of Pure and Applied Mathematics, 110*(4), 693–716.

Marzouk, Y. M., Najm, H. N., & Rahn, L. A. (2007). Stochastic spectral methods for efficient bayesian solution of inverse problems. *Journal of Computational Physics, 224*(2), 560–586. https://doi.org/10.1016/j.jcp.2006.10.010

McGrattan, E. R. (1999). Application of weighted residual methods to dynamic economic models. In R. Marimon & A. Scott (Eds.), *Computational Methods for the Study of Dynamic Economies* (pp. 114–142). Oxford and New York: Oxford University Press.

Ozen, H. C., & Bal, G. (2016). Dynamical polynomial chaos expansions and long time evolution of differential equations with random forcing. *SIAM/ASA Journal on Uncertainty Quantification, 4*(1), 609–635.

Pröhl, E. (2017). *Approximating equilibria with ex-post heterogeneity and aggregate risk*. Swiss Finance Institute: Swiss Finance Institute Research Paper Series.

Reiter, M. (2009). Solving heterogeneous-agent models by projection and perturbation. *Journal of Economic Dynamics and Control, 33*(3), 649–665.

Riesz, M. (1924). Sur le problème des moments et le théorème de Parseval correspondant. *Scandinavian Actuarial Journal, 1924*(1), 54–74. https://doi.org/10.1080/03461238.1924.10405368

Ruge-Murcia, F. J. (2007). Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics and Control, 31*(8), 2599–2636. https://doi.org/10.1016/j.jedc.2006.09.005

Scheidegger, S., & Bilionis, I. (2019). Machine learning for high-dimensional dynamic stochastic economies. *Journal of Computational Science, 33*, 68–82.

Sims, C. (2002). Solving Linear Rational Expectations Models. *Computational Economics, 20*(1), 1–20.

Soize, C., & Desceliers, C. (2010). Computational aspects for constructing realizations of polynomial chaos in high dimension. *SIAM Journal on Scientific Computing, 32*(5), 2820–2831.

Stroud, A. H. (1971). *Approximate calculation of multiple integrals*. Prentice-Hall, Englewood Cliffs, N.J.: Prentice-Hall series in automatic computation.

Szegő, G. (1939). *Orthogonal Polynomials*. American Mathematical Society.

Wiener, N. (1938). The Homogeneous Chaos. *American Journal of Mathematics, 60*(4), 897–936.

Xiu, D., & Karniadakis, G. E. (2002). The wiener-askey polynomial chaos for stochastic differential equations. *SIAM Journal on Scientific Computing, 24*(2), 619–644. https://doi.org/10.1137/s1064827501387826