BGPE Discussion Paper

No. 202

Polynomial chaos expansion: Efficient evaluation and estimation of computational models

Daniel Fehrle
Christopher Heiberger
Johannes Huber

December 2020

ISSN 1863-5733

Editor: Prof. Regina T. Riphahn, Ph.D.
Friedrich-Alexander-Universität Erlangen-Nürnberg

B G P E Discussion Paper Series

Polynomial chaos expansion: Efficient evaluation and estimation of computational models

Daniel Fehrle[a], Christopher Heiberger[b] and Johannes Huber[c]

[a]University of Augsburg, Department of Economics, Universitätsstraße 16, 86159 Augsburg, Germany, daniel.fehrle@wiwi.uni-augsburg.de
[b]University of Augsburg, Department of Economics, Universitätsstraße 16, 86159 Augsburg, Germany, christopher.heiberger@wiwi.uni-augsburg.de
[c]University of Augsburg, Department of Economics, Universitätsstraße 16, 86159 Augsburg, Germany, johannes.huber@wiwi.uni-augsburg.de

December 11, 2020

JEL classification: C11,C13,C32,C63
Keywords: Polynomial Chaos Expansion, parameter inference, parameter uncertainty, solution methods

Abstract

Polynomial chaos expansion (PCE) provides a method that enables the user to represent a quantity of interest (QoI) of a model's solution as a series expansion of uncertain model inputs, usually its parameters. Among the QoIs are the policy function, the second moments of observables, or the posterior kernel. Hence, PCE sidesteps the repeated and time consuming evaluations of the model's outcomes.

The paper discusses the suitability of PCE for computational economics. We, therefore, introduce to the theory behind PCE, analyze the convergence behavior for different elements of the solution of the standard real business cycle model as illustrative example, and check the accuracy, if standard empirical methods are applied. The results are promising, both in terms of accuracy and efficiency.

Acknowledgment: 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 initial ideas and continuing discussions.

1 INTRODUCTION

At an abstract level, computational economic models are mappings from inputs of the model to outputs of the model. The former are the model's parameters, the latter depend on the research question and comprise, e.g., the policy functions of economic agents, the second-moments of model generated time series, or the likelihood implied by a given set of observed data. The model's parameters are typically unknown and plausible values must be derived from observed data, or are even treated as random variables from the Bayesian perspective. Either way, the uncertainty of parameters translates into uncertainty regarding the model's outcomes. Estimation methods, as the generalized method of moments, the matching of impulse responses, 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 requires numerous repeated solutions of the model. Depending on the complexity of the model, estimation and sensitivity analysis can become a time consuming computational task. Polynomial chaos expansion (PCE), as employed in other scientific disciplines, offers an elegant way to deal with this problem.[1]

PCE is a method that depicts arbitrary elements of a model's solution, the quantity of interest (QoI), in terms of a series expansion of the model's parameters. Given the respective formulae, repeated evaluations are inexpensive in terms of computational time instead of repeated, potentially time-consuming, solutions of the entire model. The present paper provides a theoretical and practical primer of PCE for economists. Without limiting the applicability for other purposes, we focus on parameter estimation of dynamic stochastic general equilibrium (DSGE) models, as we are familiar with the required methods. To the best of our knowledge, applications of PCE in this context have not yet been studied in economic models.[2]

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. Given a family of stochastically independent and normally distributed random variables, which we call germs, the theorem establishes the existence of an orthogonal decomposition—with identity in the L^2 sense—of any random variable with finite second moments and measurable with respect to the germs, into Hermite polynomials in the germs. If we identify the germs with (transformations of) the model's unknown parameters, and if the model's outcome satisfies the required conditions, which apply to most computational economic models, the theory justifies an approximation of the model's outcome by a truncated series of polynomials in the unknown parameters. The so-called truncated PCE can be constructed easily from a limited number of model evaluations, and after construction of the PCE the model's outcome can be obtained uncostly by evaluation of the truncated series instead of repeated solutions of the model.

Ghanem and Spanos (1991) provide first applications of the theory to the problem of uncertain model parametrization. In such applications, we can typically restrict atten-

[1] See e.g. Kaintura et al. (2018) for a review on the increasing application of PCE in electronics and electrics.

[2] Harenberg et al. (2019) is the only study we are aware of which has studied PCE in the context of economic models so far. However, different from our work they focus on applications of PCE for structural sensitivity analysis.

tion to the far easier case with a finite number of germs. In the one-dimensional case where parametrization uncertainty is introduced by means of only one unknown parameter described by a random variable θ with (Borel) probability measure P_θ, the existence of orthogonal decompositions is the direct consequence of the property that the (orthogonal) polynomials with respect to the inner product in $L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), P_\theta)$ form a complete orthogonal system in L^2. Moreover, 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. For a finite number of unknown and stochastically independent parameters θ_i, the property also extends to tensor products of the polynomials and the product probability measure. In consequence, any L^2 mapping can be represented by a Fourier series in the orthogonal polynomials and any random variable with finite second moments which is measurable with respect to the θ_i can be written as a series of polynomials in the θ_i.

For the problem at hand, the L^2 mapping for which the Fourier series must be constructed is identified with the mapping from parameter values to the model's outcome Y. Moreover, the Fourier coefficients are defined by the inner product of this mapping with the orthogonal polynomials. If the inner product cannot be computed analytically, numerical integration rules like Gauss quadratures can be employed which, if the dimensionality of unknown parameters is not too large, require only a comparably small number of model evaluations. As the dimensionality of the problem becomes larger, sparse grid methods, such as Smolyak-Gauss quadrature can help or, alternatively, the coefficients can be obtained from least squares.

After construction of the truncated PCE, it can be used for inexpensive evaluations of the model outcome. First, statistical properties of the model outcome can be derived directly from the PCE and the parameters' distributions. The statistical properties can then be used to quantify the effects of parameter uncertainty. For example, the variance of the model outcome can be used as a first indicator for a sensitivity analysis. Moreover, Harenberg et al. (2019) propose a sensitivity analysis on the basis of Sobol' indices which can be obtained directly from the PCE. The analysis additionally provides necessary conditions for parameter identification in structural estimations. Second, the Fourier expansion can also be used as a pointwise approximation for the mapping between parameters and the QoI. Thus, estimation methods which require repeated recalculations of the model outcome can be sped up significantly. Since Bayesian inference naturally combines the specification of a-priori parameter uncertainty in form of prior distributions with the necessity for repeated model solutions, it provides an especially well-suited setting for the implementation of PCE. 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 in economic models.

We apply the method of PCE to the benchmark real business cycle (RBC) model, since this model is suited as illustrative example due to its well-known and simplistic nature. We analyze the convergence behaviour of the PCE—in the sense of the L^2 norm of the approximation error over the parameters' support—as the degree of truncation is increased. Our

analysis starts with an example where three parameters are assumed unknown, namely the capital share in production, the coefficient of relative risk aversion and the autocorrelation parameter of total factor productivity, and considers the PCEs of various model outcomes including the model's linear solution, a projection solution, the variables' second moments and the impulse response function. Although we assume rather "loose" distributions for the unknown parameters, we find linear convergence speed in all cases, and remarkably well approximations can be obtained already with a rather small degree of truncation and a small number of model evaluations. If the model outcome, e.g the linearized policy function, has to be evaluated for a sample of 100,000 parameter values, the PCE with truncation degree 7 provides an approximation with L^2 error of 10^{-3} while the computational time for construction and evaluation is lower by the factor 30 compared to repeated computations.

We extend our example to the higher-dimensional problem where all six model parameters are assumed unknown. Compared to full-grid quadrature rules, sparse-grid quadrature rules and least squares provide less accurate derivations of the PCE coefficients. In consequence, the approximated PCEs require a higher degree of truncation in order to deliver the same accuracy. However, they also require significantly less time for construction. A comparison of computational time versus the approximation's accuracy shows that the PCE constructed from sparse-grid quadrature is most efficient followed by least squares. Yet, for higher degrees of truncation, inaccuracies in the PCE coefficients derived from least squares eventually become dominant and even reverse convergence.

Our analysis continues with Monte Carlo experiments as in Ruge-Murcia (2007) where we gauge the quality of the model's PCE when used for several empirical methods. More specifically, we estimate the model's parameters by generalized method of moments (GMM), simulated method of moments (SMM), maximum-likelihood estimation (MLE) and Bayesian estimation (BE) but use PCE to evaluate the QoI for different parameter values. Compared to the benchmark procedure of repeated solutions, we find that 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.

The remainder of the paper is structured as follows. First, we give a simple example to outline the concept of PCE in section 2. In section 3 we review the basic theory for the existence of polynomial chaos expansions and present the most common practical methods to compute the PCE coefficients. Section 4 discusses different applications of the PCE, either to evaluate statistical properties of the model outcome or for pointwise approximation of the mapping from the parameters to the model outcome. We particularly highlight its application to construct surrogates for the model's linear solution or for projection solutions and to approximate gradients. In section 5, we apply the method to the benchmark RBC model and discuss the basic results and potential drawbacks. Section 6 concludes. More detailed derivations are found in the appendix. MATLAB® code is available from the authors upon request.

2 A Simple Example

Before introducing the theoretical framework of PCE, we first want to outline the concept at hand of a simple example. 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[3]

$$h_{ij}(\psi(s)) = \sum_{n=0}^{\infty} \hat{h}_{ij}^{(n)} L_n(s). \tag{1}$$

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

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

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[4]

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

Table 1 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

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

[4] 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

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

Note that this expression is identical to the more general form in (6).

Table 1: 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

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.,[5]

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

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 (1), i.e.

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

in order to (approximately) evaluate $h_{11}(\vartheta)$ for arbitrary parameter values without further need of direct numerical evaluations.[6] Figure 1 shows a comparison between 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.,

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

[5] For comparison, exact integration yields

$$\hat{h}_{11}^{(0)} = \frac{1}{2}\int_{-1}^{1} \frac{-1}{1-\frac{(s+1)b}{2}}\,\mathrm{d}s = \frac{\ln(1-b)}{b} = -2.56,\ \hat{h}_{11}^{(1)} = \frac{3}{2}\int_{-1}^{1} \frac{-s}{1-\frac{(s+1)b}{2}}\,\mathrm{d}s = \frac{6-3b}{b^2}\ln(1-b) + \frac{6}{b} = 2.71.$$

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

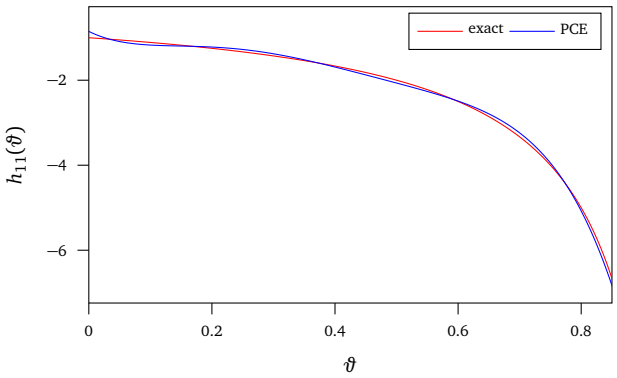


Figure 1: Example: Exact Evaluation and PCE (numerical integration)

3 Generalized Polynomial Chaos Expansions

We begin by reviewing the basic idea and theory behind the concept of PCE. While PCE proved useful for a variety of applications, we focus on their implementation to efficiently evaluate computationally expensive model outcomes when one or more of the model's inputs, e.g. model parameters, are uncertain.

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.[7] This mapping between the ϑ_i and y is described by

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

where $h \colon \Theta \to \mathbb{R}^m, \Theta = \times_{i=1}^{k} \Theta_i \subset \mathbb{R}$. 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 describing them by appropriately distributed random variables. Therefore, let $(\Omega, \mathscr{A}, P)$ denote a sufficiently rich probability space so that any uncertain model input parameter can be described by some real valued random variable $\theta_i \colon \Omega \to \mathbb{R}, i = 1, \dots, k$, where the real line is equipped with the Borel sigma-algebra $\mathscr{B}(\mathbb{R})$. Moreover, let $\xi_1, \dots, \xi_k$ 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

[7] 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.

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 $\xi_1, \dots, \xi_k$ cover the same stochastic information as the uncertain model parameters, i.e.

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

   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 $i = 1, \dots, k$ and $n \in \mathbb{N}_0$.

Moreover, we write $\boldsymbol{\theta} := (\theta_1, \dots, \theta_k) \colon \Omega \to \mathbb{R}^k$ and $\boldsymbol{\xi} := (\xi_1, \dots, \xi_k) \colon \Omega \to \mathbb{R}^k$ 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 $\mathscr{B}(\mathbb{R}^k)$. For each $i = 1, \dots, k$, let $P_{\xi_i} := P \circ \xi_i^{-1}$ denote the probability measure of ξ_i on $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 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 $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k))$. The Hilbert space (of equivalence classes) of square integrable real valued functions on $(\mathbb{R}, \mathscr{B}(\mathbb{R}), P_{\xi_i})$ is denoted by

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

where the inner product is defined by

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

We use the notation $\| \cdot \|_{L_i^2}$ for the induced norm on L_i^2. We introduce the analogous notation, i.e. $L^2 := L^2(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), \mathrm{d}P_{\boldsymbol{\xi}})$, for the space of square integrable real valued functions on $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), P_{\boldsymbol{\xi}})$ 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 $w_i \colon \mathbb{R} \to \mathbb{R}_+$, 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

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

so that $L_i^2 = L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), w_i(s)\,\mathrm{d}s)$ and $L^2 = L^2(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), w(s)\,\mathrm{d}s)$ 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 $\psi_i \colon \mathbb{R}^k \to \mathbb{R}$ which satisfy

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

We write $\psi := (\psi_1, \dots, \psi_k) \colon \mathbb{R}^k \to \mathbb{R}^k$ so that $\boldsymbol{\theta} = \psi \circ \boldsymbol{\xi}$. Moreover, note that $\sigma(\boldsymbol{\xi}) = \sigma(\boldsymbol{\theta})$ also implies the existence of a measurable, inverse mapping ψ^{-1} with $\psi \circ \psi^{-1} = \psi^{-1} \circ \psi = \mathrm{id}$. 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 $(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), P_{\boldsymbol{\xi}})$, i.e. $\psi_i \in L^2$ for all $i = 1, \dots, k$.[8]

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 $Y \colon \Omega \to \mathbb{R}$. 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.

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

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

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

3.1 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

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

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 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 2 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 dis-

Table 2: 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	$\theta \sim \mathrm{Gamma}(\alpha, \beta)$[a]	$\xi := \beta\theta$	$\psi(s) = \frac{s}{\beta}$	General Laguerre $La_n^{(\alpha-1)}$
Inverse Gamma	$\theta \sim \text{Inv-Gamma}(\alpha, \beta)$[a]	$\xi := \frac{\beta}{\theta}$	$\psi(s) = \frac{\beta}{s}$	General Laguerre $La_n^{(\alpha-1)}$

[a] We use the scale-rate notation.

tributed. More details for these classes are given in Appendix A. In all of the cases presented in Table 2 the respective families of orthogonal polynomials $\{q_n\}_{n\in\mathbb{N}_0}$ form a complete orthogonal system, i.e. lie densely in $L^2 = L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), P_\xi) = L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), w(s)\,\mathrm{d}s)$ where w is the corresponding probability density of ξ.[9] 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 $(\mathbb{R}, \mathscr{B}(\mathbb{R}))$ 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

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

Note that identity and convergence is understood in L^2 which also implies pointwise

[8] 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.

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

convergence a.e. for a subsequence but not pointwise convergence.[10] Moreover, since $P_\theta = P_\xi \circ \psi^{-1}$, also $h = \sum_{n=0}^{\infty} \hat{y}_n (q_n \circ \psi^{-1})$ in $L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), P_\theta)$.

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

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

$$Y = h(\theta) = h(\psi(\xi)) = \sum_{n=0}^{\infty} \hat{y}_n q_n(\xi) \text{ in } L^2(\Omega, \mathscr{A}, P). \tag{3}$$

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

$$\hat{\vartheta}_n = \|q_n\|_{L^2}^{-2} \langle \psi, q_n \rangle_{L^2} = \|q_n\|_{L^2}^{-2} \int_{\mathbb{R}} \psi q_n \, \mathrm{d}P_\xi, \tag{4}$$

$$\hat{y}_n = \|q_n\|_{L^2}^{-2} \langle h \circ \psi, q_n \rangle_{L^2} = \|q_n\|_{L^2}^{-2} \int_{\mathbb{R}} (h \circ \psi) q_n \, \mathrm{d}P_\xi. \tag{5}$$

Now in practice, equations (2)-(3) 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 (4)-(5) 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.

3.1.1 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 principal any choice of ψ which satisfies Assumption 2 is possible. One could then construct the family of

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

orthogonal polynomials from the germ's distribution and the expansion coefficients could be derived by numerical integration of (4) 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 2. In this case the expansion (2) 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 way that it is compatible to 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[11]

$$\xi \coloneqq F_\xi^{-1} \circ F_\theta \circ \theta$$

yields the desired distribution for ξ. Conversely,

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

and the expansion coefficients can again be computed from (4) by numerical integration.

3.1.2 Polynomial Chaos Expansion of the Model Outcome

While the expansion of the model parameter can be directly controlled by the appropriate choice of the germ, the expansion of the model outcome of interest requires some evaluations of the model.

Spectral Projection The first approach derives the polynomial chaos coefficients $\hat{y}_n$ by applying numerical integration methods to (5). For example, if ξ possesses a probability density function w, then (5) 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

$$\hat{y}_n \approx \|q_n\|_{L^2}^{-2} \sum_{i=1}^{M} h(\psi(s_i)) q_n(s_i) \omega_i \approx \|q_n\|_{L^2}^{-2} \sum_{i=1}^{M} h\left(\sum_{m=1}^{N} \hat{\vartheta}_m q_m(s_i) \right) q_n(s_i) \omega_i, \tag{6}$$

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 $h(\psi(s_i)) \approx$

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

$h\left(\sum_{m=1}^{N} \hat{\vartheta}_m q_m(s_i)\right)\big)$ at each of the M nodes. Since the quadrature rule with M nodes 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 $n = 1, \ldots, N$. 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 $s_j, j = 1, \ldots, M$, from the distribution P_ξ or select them according to regression design principles. After computing the corresponding model outcomes $Y_j = h(\psi(s_j)) \approx h\left(\sum_{m=1}^{N} \hat{\vartheta}_m q_m(s_j)\right)\big)$ the expansion coefficients are determined from

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

The number of 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 $M = 3(N+1)$.

Stochastic Galerkin For both methods discussed in the preceding paragraphs, the computation of the expansion coefficients is detached from the underlying procedure from which the model outcome is computed. This is different for 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

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

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[12], say $\Phi_j, j =$

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

$1, \ldots, d$, 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 serve as our model outcome of interest and should not be confused with the Fourier coefficients of the PCE, one can, for example, select d appropriate collocation points $x_1, \ldots, x_d \in S$ and solve the non-linear system of equations given by

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

for $y_1, \ldots, y_d$.

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

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

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

$$R\left(\sum_{j=1}^{d} Y_j \Phi_j, x_i; \theta\right) = 0 \text{ for all } i = 1, \ldots, d. \tag{7}$$

In order to avoid the necessity for repeated and potentially computational 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[13]

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

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

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

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

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 (7) with their PCE and impute the corresponding conditions

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

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

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

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 3.1.1.

3.2 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 $i = 1,\dots,k$ 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

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

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

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

Since stochastic independence of the ξ_i implies that $P_\xi = \otimes_{i=1}^k P_{\xi_i}$, 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 $i = 1, \dots, k$ 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 2 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 $\{\hat{\vartheta}_{i\alpha}\}_{\alpha\in\mathbb{N}_0^k} \subset \mathbb{R}, i = 1, \dots, k$, and $\{\hat{y}_\alpha\}_{\alpha\in\mathbb{N}_0^k} \subset \mathbb{R}$ such that

$$\psi_i = \sum_{\alpha\in\mathbb{N}_0^k} \hat{\vartheta}_{i\alpha} q_n \text{ in } L^2 = L^2(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), P_\xi), \tag{8}$$

$$h \circ \psi = \sum_{\alpha\in\mathbb{N}_0^k} \hat{y}_\alpha q_\alpha \text{ in } L^2 = L^2(\mathbb{R}^k, \mathscr{B}(\mathbb{R}^k), P_\xi). \tag{9}$$

The second expansion can again be written equivalently as

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

Therefore, the parameters θ_i and the model outcome Y are again representable in L^2 by a PCE in the germs ξ through

$$\theta_i = \psi_i \circ \xi = \sum_{\alpha\in\mathbb{N}_0^k} \hat{\vartheta}_{i\alpha} q_\alpha(\xi) \text{ in } L^2(\Omega, \mathscr{A}, P), \tag{10}$$

$$Y = h \circ \boldsymbol{\theta} = h \circ \psi \circ \xi = \sum_{\alpha\in\mathbb{N}_0^k} \hat{y}_\alpha q_\alpha(\xi) \text{ in } L^2(\Omega, \mathscr{A}, P). \tag{11}$$

Moreover, the expansion coefficients are determined by

$$\hat{\vartheta}_{i\alpha} = \|q_\alpha\|_{L^2}^{-2} \langle \psi_i, q_\alpha \rangle_{L^2} = \|q_\alpha\|_{L^2}^{-2} \int_{\mathbb{R}^k} \psi_i q_\alpha \, \mathrm{d}P_\xi, \tag{12}$$

$$\hat{y}_\alpha = \|q_\alpha\|_{L^2}^{-2} \langle h \circ \psi, q_\alpha \rangle_{L^2} = \|q_\alpha\|_{L^2}^{-2} \int_{\mathbb{R}^k} (h \circ \psi) q_\alpha \, \mathrm{d}P_\xi, \tag{13}$$

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

Equations (12)-(13) 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

$$S_N^{\text{tot}}(\theta_i) = S_N^{\text{tot}}(\psi_i \circ \xi) := \sum_{\alpha \in \mathbb{N}_0^k, |\alpha| \leq N} \hat{\vartheta}_{i\alpha} q_\alpha(\xi),$$
$$S_N^{\text{tot}}(Y) = S_N^{\text{tot}}(h \circ \psi \circ \xi) := \sum_{\alpha \in \mathbb{N}_0^k, |\alpha| \leq N} \hat{y}_\alpha q_\alpha(\xi),$$

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

$$S_N^{\max}(\theta_i) = S_N^{\max}(\psi_i \circ \xi) := \sum_{\alpha \in \mathbb{N}_0^k, \|\alpha\|_\infty \leq N} \hat{\vartheta}_{i\alpha} q_\alpha(\xi),$$
$$S_N^{\max}(Y) = S_N^{\max}(h \circ \psi \circ \xi) := \sum_{\alpha \in \mathbb{N}_0^k, \|\alpha\|_\infty \leq N} \hat{y}_\alpha q_\alpha(\xi),$$

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

In order to compute the expansion coefficients from their defining equations (12)-(13), it is straightforward to adapt the methods from section 3.1.2 to the multidimensional case. However, this typically introduces the curse of dimensionality.

First, this issue becomes particularly problematic if the integrals are computed by Gauss-quadrature rules. 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 (13) can be well approximated by multivariate polynomials which rise up 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 (13) 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, as e.g. Smolyak-Gauss quadrature which is illustrated in Appendix B and analyzed in the numerical example in section 5, can help to reduce the computational effort that is required for similar integration quality.

Second, the burden of higher-dimensional parameter vectors also appears in similar form if the PCE coefficients are determined by least squares. However, while the number of coefficients which 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.

4 Applications of Generalized Polynomial Chaos Expansions

After its construction, the PCE of the model outcome can be used for computational inexpensive evaluations of the model. On the one hand, statistical properties of the model

outcome, as induced by the predefined distribution of the uncertain input parameters, can be derived directly from the PCE. On the other hand, the expansion can also be used as a pointwise approximation of the model outcome for different parameter values.

Evaluation of Statistical Properties Convergence in $L^2(\Omega, \mathcal{A}, P)$ of the series expansion in (11) 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 $\mathbf{0} := (0, \dots, 0) \in \mathbb{N}_0^k$) yields

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

and

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

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.

Using the Expansion as Pointwise Approximation for the Model Outcome A truncated version of the Fourier series expansion (9) can also be used as a pointwise approximation for the mapping h between model input parameters and model outcome

$$h(\vartheta) \approx S_N(h \circ \psi)(\psi^{-1}(\vartheta)) = \sum_{\alpha \in \mathbb{N}_0^k, |\alpha| \leq N} \hat{y}_\alpha q_\alpha(\psi^{-1}(\vartheta)). \tag{14}$$

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 a.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 $L^2(\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k), P_\theta)$.[14] In the sense of the induced metric, it is therefore the best approximation of h by multivariate

[14] 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.

polynomials of total degree up to N, i.e. it minimizes the mean-squared error over the support of $P_{\boldsymbol{\theta}}$.

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 $g(.;\vartheta)\colon S \to \mathbb{R}^{n_x+n_y}$ so that

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

If, for any arbitrary $x \in S$ and for 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

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

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

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

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

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 $\hat{A}_\alpha$ coincide 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 meth-

ods.[15] 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

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

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 $\vartheta \mapsto c_i(\vartheta)$.

Surrogate for Gradients The truncated PCE in (14) 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

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

This approximation can be useful if such derivatives must be evaluated at a potential 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 with respect to the parameters.

5 Numerical Analysis

In this section we present the numerical implementation of a PCE for a benchmark RBC model. First, we analyze the convergence behaviour of the series expansion for different model outcomes of interest. More specifically, the model outcomes considered include the linear solution, the second moments and the impulse response functions of the model's

[15] See, for instance Judd (1996), Chapter 11, Heer and Maussner (2009), Chapter 6, Judd (1992) or McGrattan (1999).

variables to a one time shock—both computed from the model's linear policy—as well as a global projection solution. Moreover, we compare different methods to compute the PCE coefficients in terms of accuracy and efficiency. Finally, we perform Monte-Carlo experiments where we evaluate the performance of the PCE for empirical applications as matching moments and likelihood-based approaches.

5.1 The model

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

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

where Y_t, C_t, N_t, I_t and K_t denote output, consumption, working hours, investment 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

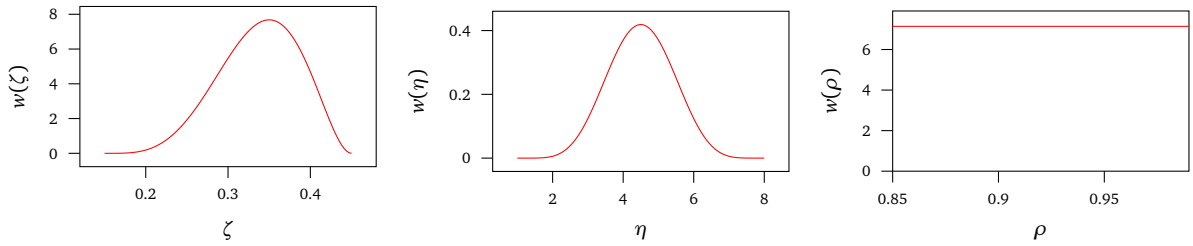
$$x_t := \begin{pmatrix} K_t \\ z_t \end{pmatrix} \quad \text{and} \quad y_t := \begin{pmatrix} Y_t \\ C_t \\ N_t \\ I_t \end{pmatrix}.$$

5.2 Convergence Behaviour

First, in order to study the basic convergence behaviour 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 Figure 2. The transformations ψ_i between unknown parameters and germs are fixed as in Table 2 and the remaining parameters are calibrated as summarized in Table 3.



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

Figure 2: Distributions of uncertain parameters

Table 3: Calibration I

Parameter	Description	Value
β	discount factor	0.994
δ	rate of capital depreciation	0.014
N	steady state labor supply	0.3
σ	standard deviation of innovations to log TFP	0.01

Linear Policy Function The first model outcome which 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 $A(\vartheta) = \big(a_{ij}(\vartheta)\big)_{\substack{i=1,\dots,6 \\ j=1,2}} \in \mathbb{R}^{6\times 2}$ can be easily computed numerically from available methods. As described in section 4, 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 $a_{ij}\colon \vartheta \mapsto a_{ij}(\vartheta)$ the truncated PCE[16]

$$a_{ij}^{(N)}(\vartheta) := S_N^{\text{tot}}(a_{ij} \circ \psi)(\psi^{-1}(\vartheta)) = \sum_{\alpha\in\mathbb{N}_0^3, |\alpha|\leq N} \hat{a}_{ij\alpha} q_\alpha(\psi^{-1}(\vartheta)). \tag{15}$$

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 behaviour 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 integration errors in (11) (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

[16] 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.

quadrature rules and determine the L^2 error from

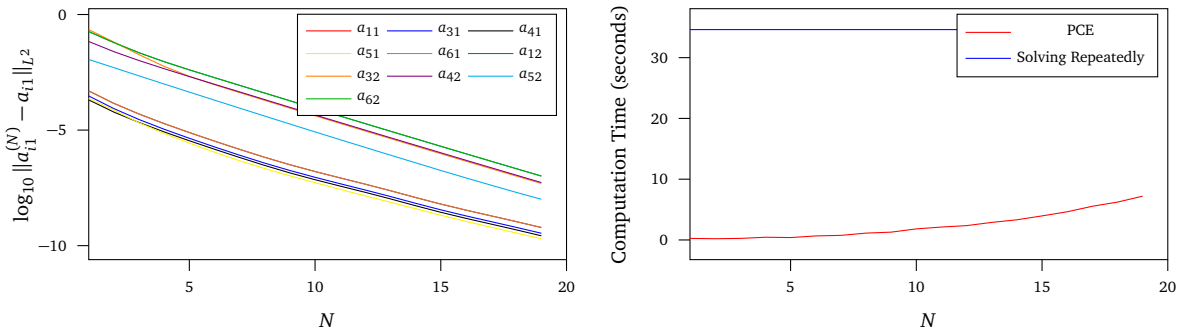
$$\|a_{ij}^{(N)} - a_{ij}\|_{L^2} = \left(\int_{\mathbb{R}^3} \left(a_{ij}^{(N)}(\vartheta) - a_{ij}(\vartheta) \right)^2 dP_\theta \right)^{1/2} \approx \left(\frac{1}{M} \sum_{i=1}^{M} \left(a_{ij}^{(N)}(\vartheta^{(i)}) - a_{ij}(\vartheta^{(i)}) \right)^2 \right)^{1/2} \tag{16}$$

where we draw $M = 10^5$ iid sample points $\vartheta^{(i)}$ from the distribution of θ. The results are presented in Figure 3(a) 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, Figure 3(b) 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 which 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 outcome 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 to subsequently use this PCE of the linear policy to compute the second moments.

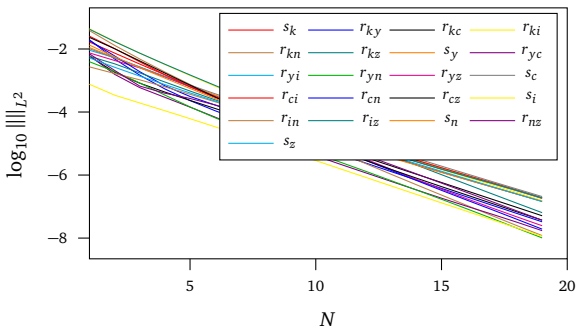
Figure 3(c) again shows linear convergence of the PCEs for each second moment. The L^2 error in the approximation of the model's moments has fallen to the order of magnitude of -3 by $N = 7$ and further declines to -6 by $N = 19$. Moreover, the computation time of the PCE versus the time for repeated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

Impulse Response Function The next model outcome 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},

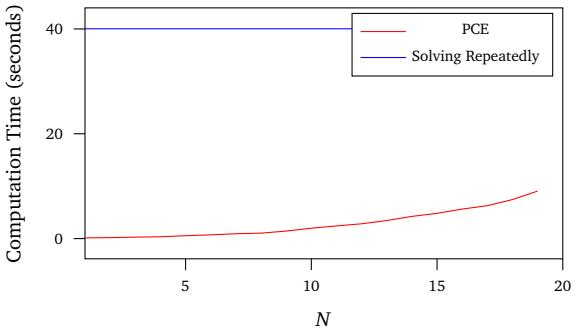


(a) L^2 convergence linear policy

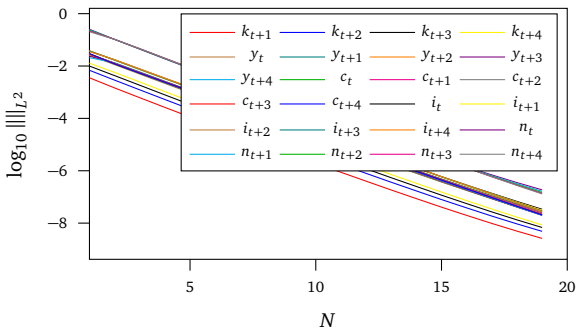
(b) Computation Time linear policy



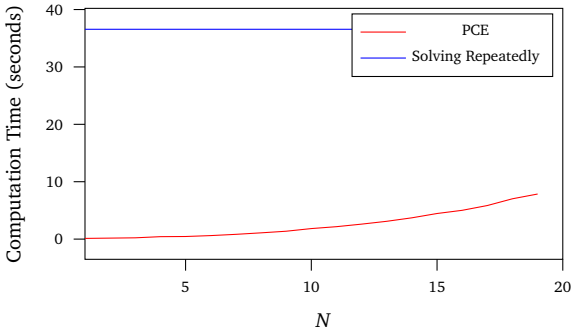
(c) L^2 convergence second moments



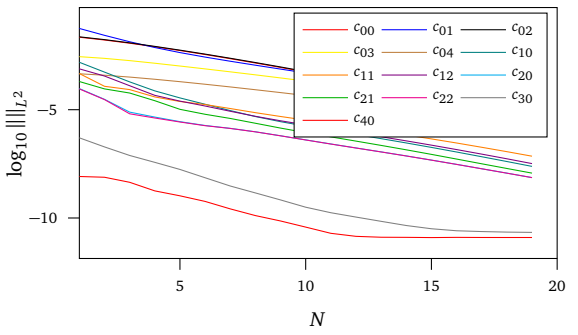
(d) Computation Time second moments



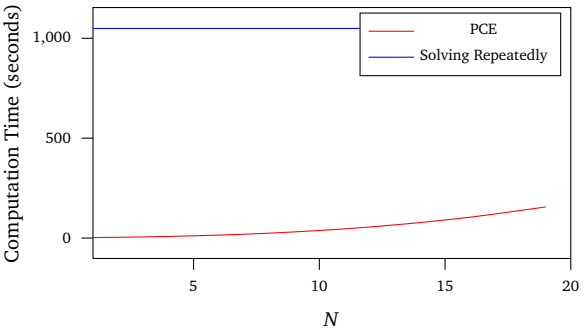
(e) L^2 convergence IRF



(f) Computation Time IRF



(g) L^2 convergence global solution



(h) Computation Time global solution

Figure 3: L^2 convergence of PCE and computation time on an Intel® Core™i7-7700 CPU @ 3.60GHz

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 Figure 3(e). 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 throughout 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

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

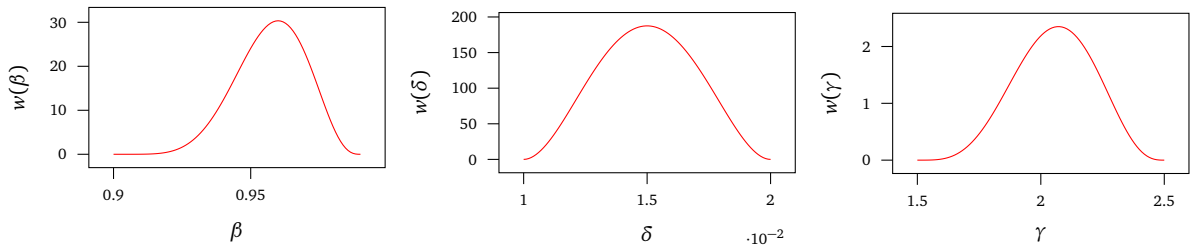
where we further introduce the transformation $n_t := \ln(N_t/(1-N_t))$. The T_i are Chebyshev polynomials of degree i and $[\underline{k}; \bar{k}] \times [\underline{z}; \bar{z}] = [\ln(0.8); -\ln(0.8)] \times [-3\frac{\sigma}{\sqrt{1-\rho^2}}; 3\frac{\sigma}{\sqrt{1-\rho^2}}]$ is the domain of the 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 such way that the model's Euler equation holds exactly at 13 appropriately selected collocation points.[17]

We discussed in section 4 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 Figure 3(g) and the time for construction and evaluation of the PCEs in Figure 3(h) remains throughout significantly smaller than the time for repeated computations of the global solution.

5.3 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 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

[17] 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

Figure 4: Distributions of uncertain parameters

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 additionally 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. Sparse-grid methods as well as least squares give fundamentals for a rising number of more efficient alternatives. Kaintura et al. (2018) and Harenberg et al. (2019) give a short discussion.

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 $\theta := \begin{pmatrix} \zeta & \eta & \rho & \beta & \delta & \gamma \end{pmatrix}$.[18] The assumed distributions for ζ, η and ρ remain as before in Figure 2 and the distributions of the additional unknown parameters are chosen as

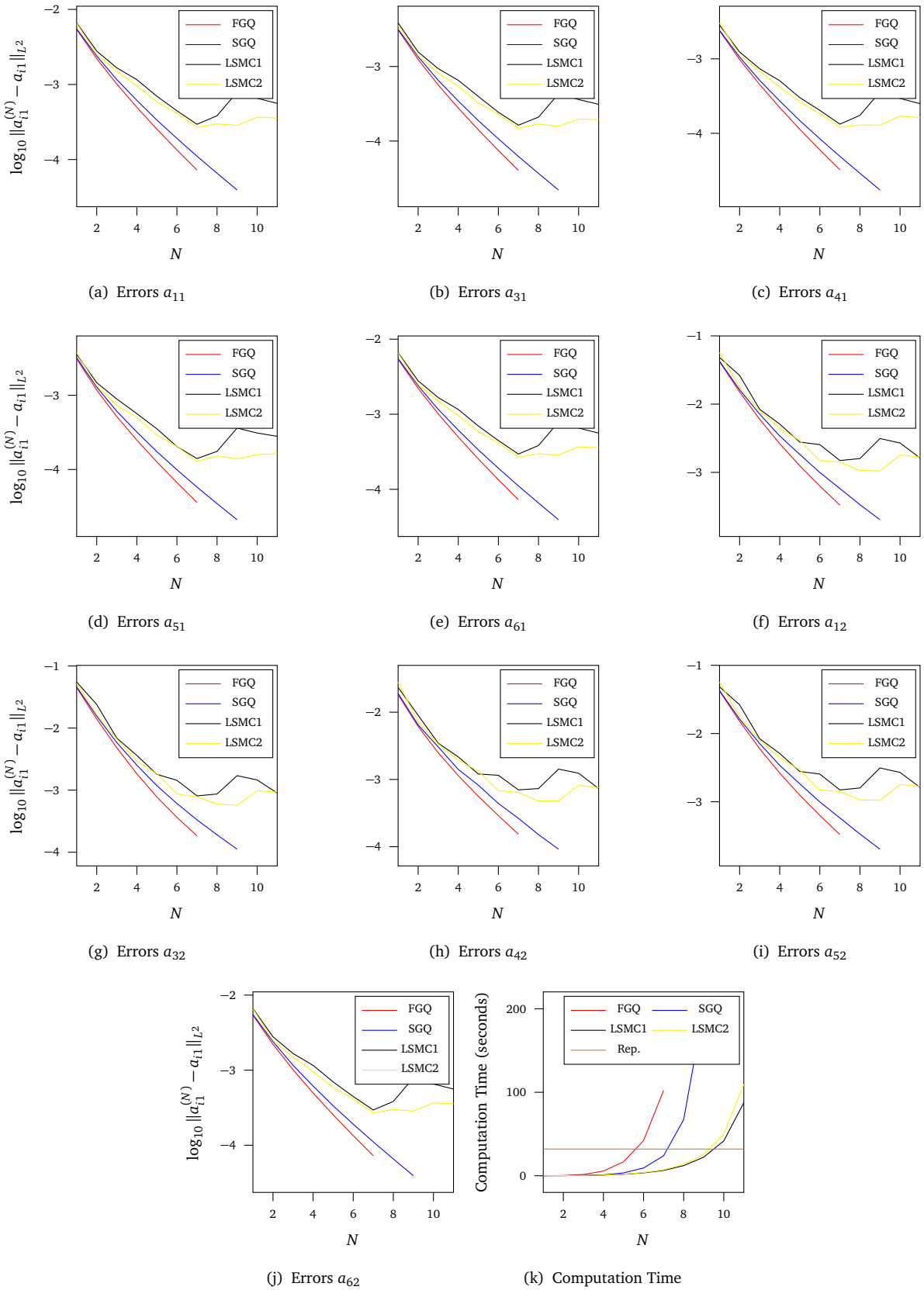
$$\beta \sim 0.9 + 0.09 \cdot \text{Beta}(7,4),\ \delta \sim 0.01 + 0.01 \cdot \text{Beta}(3,3),\ \gamma \sim 1.5 + 1 \cdot \text{Beta}(5,4).$$

The probability densities for β, δ and γ are visualized in Figure 4.

We compute the truncated PCE (15) for each mapping $a_{ij} : \vartheta \mapsto a_{ij}(\vartheta)$ in the linear policy $A(\vartheta) = \left(a_{ij}(\vartheta)\right)_{\substack{i=1,\dots,6 \\ j=1,2}} \in \mathbb{R}^{6\times 2}$. The PCE coefficients are now determined either by i) a full-grid Gauss quadrature rule with $N+1$ nodes for each parameter (FGQ), ii) a sparse-grid Smolyak-Gauss quadrature rule with linear growth where the level is set such 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 (16) from a draw of $M = 10^5$ iid sample points from the parameter's distribution.

Figure 5 shows the convergence of the truncated (approximated) PCEs with approximated coefficients for increasing N. As expected, the PCE constructed from a full-grid

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



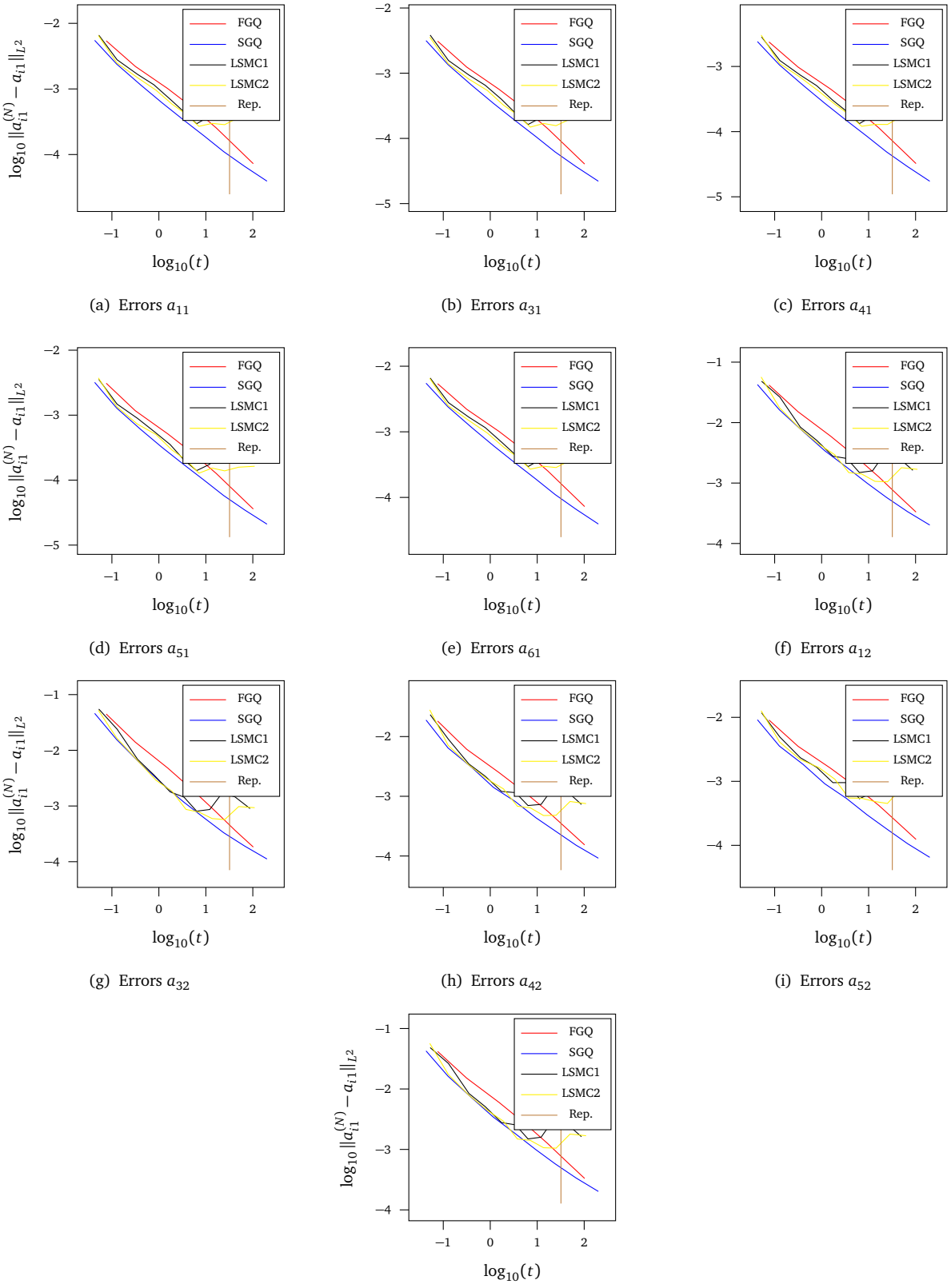
(a) Errors a_{11} (b) Errors a_{31} (c) Errors a_{41}

(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

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



(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}

Figure 6: L^2 Convergence of PCE with approximated coefficients and computation Time on an Intel® Core™i7-7700 CPU @ 3.60GHz

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. In fact, since inaccuracies in the coefficients of higher degree polynomials may have large impact on the L^2 error of the PCE,[19] 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 5(k) 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, Figure 6 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 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.

5.4 Monte Carlo experiments for empirical methods

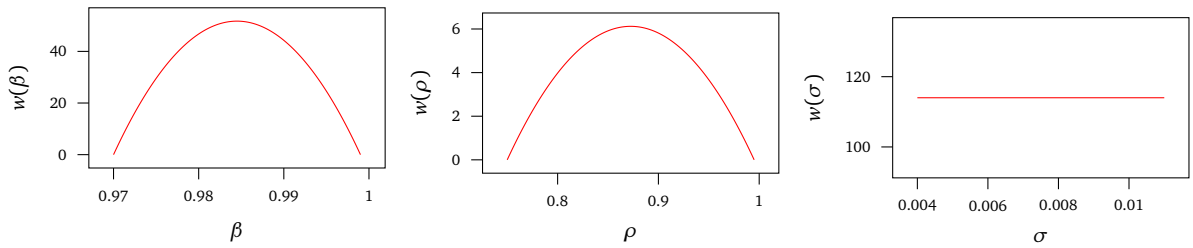
Design Our Monte Carlo study follows Ruge-Murcia (2007) and analyzes the performance of PCE when applied to different estimation methods. We set the vector of uncertain parameters to $\theta := (\beta, \rho, \sigma)$ and choose the following probability distributions with support $\Theta := [0.97; 0.999]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

$$\beta \sim 0.97 + 0.029 \cdot \text{Beta}(2,2),\ \rho \sim 0.75 + 0.245 \cdot \text{Beta}(2,2),\ \sigma \sim 0.004 + 0.008 \cdot U(0,1).$$

Figure 7 illustrates the uncertain parameters' probability densities and the remaining parameters are calibrated as summarized in Table 4.

Matching Moments To estimate the parameters by matching moments, we choose the following 5 targets: i) the variance of output and of working hours, ii) the autocovariance (lag 1) of output and of 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 in order to evaluate the model's second

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



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

Figure 7: Distributions of uncertain parameters

Table 4: Calibration of the model

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	$\sigma \sim 0.004 + 0.008 \cdot U(0,1)$

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 or iii) constructing the PCE of the model's second moments which we then evaluate. 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 a sufficient accuracy to demonstrate the capabilities.[20] 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.

Table 5 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 well approximation

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

Table 5: Monte Carlo Results - GMM

Benchmark (Repeated Solution)			
	Total average		
Time:	00:01.25		
PCE Policy Function			
	Total average	PCE	Estimation average
Time:	00:00.5	00:00.05	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			
	Total average	PCE	Estimation average
Time:	00:03.44	00:03.11	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

Notes: 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.

Table 6: Monte Carlo Results - SMM

Benchmark (Repeated Solution)			
	Total average		
Time:	01:12.82		
PCE Policy Function			
	Total average	PCE	Estimation average
Time:	00:34.67	00:00.03	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			
	Total average	PCE	Estimation average
Time:	00:58.03	00:57.73	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

Notes: 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 quadratur level of the expanded policy function is 7 and of the second moments 13.

which results in deviations from the benchmark mostly smaller than one permille in comparison to the range of the parameter's distribution. Estimation errors rise if the model's second moments are directly approximated by PCE. However, the average relative errors remain below two permille for all parameters and is 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 in order to derive the model's moments. However, the computation 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 6.

We find again that the policy function's PCE provides a remarkably well 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 PCE once the expansion of the QoI is calculated.

Likelihood-based Estimation We proceed to analyze the performance of PCE in MLE and in 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.[21] 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 or iii) constructing the PCE of the likelihood which we then evaluate. In order to avoid problems with weak identification and in order to focus on the quality of PCE in the estimation procedure, MLE is unusually applied to data in levels instead of the relative deviation from steady state.

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

For BE the priors remain the same as in Table 4. 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. 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.[22] 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

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

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

Table 7 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 from the benchmark version remain remarkably small. The average error concerning the policy function's PCE estimation is smaller than one permille in comparison 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 the 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 8 summarizes the results from the PCE based methods—approximation of the policy function or of 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 then 1.25 percent. Recognizing the fact 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.[23]

Discussion Our study of PCE in estimation of a standard real business cycle (RBC) model shows that the PCE based methods can accurately reproduce the same results as the bench-

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

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

Table 7: Monte Carlo Results - Maximum Likelihood Estimation

Benchmark (Repeated Solution)			
	Total average		
Time:	00:20.60		
PCE Policy Function			
	Total average	PCE	Estimation average
Time:	00:18.79	00:00.20	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			
	Total average	PCE	Estimation average
Time:	00:10.81	00:10.36	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

Notes: 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.

mark method of 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, we show the gains in efficiency are almost 99 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 in order to achieve compactness of the support. This procedure is unconventional in Bayesian estimation of DSGE Models but helps for PCE. First and foremost, 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 be computed in a numerically stable way at all nodes of the quadrature rules.[24]

[24]For example, larger values of the capital share quickly result in numerical problems for the computation

Table 8: Monte Carlo Results - Bayesian Estimation

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		
j	x:	Mean:	Quantile: 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
PCE Posterior-Kernel									
Time:		Total average 00:16.82			PCE 00:11.00		Estimation average 00:05.82		
j	x:	Mean:	Quantile: 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

Notes: 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 quadratur level of the expanded policy function is 9 and of the second moments 13.

In non-Bayesian approaches, the application of PCE demands the otherwise not necessary specification of prior distributions. Moreover, L^2 convergence of the series expansion is achieved w.r.t. this prior distribution of the parameters. Estimation fails if the true parameter value is at odds to the choice of priors.

Similarly, Lu et al. (2015) show that the use of 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 is an indicator for a sufficient high truncation level.

Finally, the success of PCE is determined by the ratio of the number of model evaluations necessary in order to compute the coefficients and the number of model evaluations in the estimation method. Hence, PCE works best in cases with a small number of unknown parameters where estimation demands many model evaluations, but PCE loses efficiency in higher dimensional problems.

6 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 evaluate statistical properties of the QoI from the PCE and how to use the expansion as a pointwise approximation for the QoI. Further, surrogates for a linearized policy function, for a policy function based on projection methods, and for gradients of the model's QoI are presented.

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. For the higher dimensional problem with six unknown parameters, sparse-grid quadrature is the most efficient method compared to least squares and a full-grid quadrature. Monte Carlo experiments for different empirical methods show that the PCE based methods can accurately reproduce the same results as 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 critically suffers from the curse of dimensions in problems with a large number of unknown parameters. Further, poorly chosen priors may affect the accuracy of the estimates.

Despite of these potential drawbacks, PCE is a powerful tool for a broad set of applications. We hope that the article can encourage applications of PCE in economics, especially for parameter inference in complex models where standard techniques are infeasible.

of the linear approximation of policy function.

REFERENCES

O. J. Blanchard and C. M. Kahn. The Solution of Linear Difference Models under Rational Expectations. *Econometrica*, 48(5):1305–1311, July 1980.

R. H. Cameron and W. T. Martin. The Orthogonal Development of Non-Linear Functionals in Series of Fourier-Hermite Functionals. *Annals of Mathematics*, 48(2):385–392, Apr. 1947.

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

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

B. Heer and A. Maussner. *Dynamic General Equilibrium Modeling: Computational Methods and Applications*. Springer Berlin Heidelberg, 2009. ISBN 9783540856856. URL https://books.google.de/books?id=ZdlEAAAAQBAJ.

E. P. Herbst and F. Schorfheide. *Bayesian estimation of DSGE models*. The Econometric and Tinbergen Institutes lectures. Princeton University Press, Princeton ; Oxford, 2016. ISBN 9780691161082.

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

D. Jackson. *Fourier Series and Orthogonal Polynomials*. Mathematical Association of America, 1941.

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

K. L. Judd. Approximation, perturbation, and projection methods in economic analysis. In H. M. Amman, D. A. Kendrick, and J. Rust, editors, *Handbook of Computational Economics*, volume 1 of *Handbook of Computational Economics*, chapter 12, pages 509–585. Elsevier, 1996.

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

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

F. Lu, M. Morzfeld, X. Tu, and A. J. Chorin. Limitations of polynomial chaos expansions in the bayesian solution of inverse problems. *Journal of Computational Physics*, 282: 138–147, feb 2015. doi: 10.1016/j.jcp.2014.11.010.

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

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

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

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

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

G. Szegő. *Orthogonal Polynomials*. American Mathematical Society„ 1939.

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

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

APPENDIX
Polynomial chaos expansion: Efficient evaluation and estimation of computational models

A Orthogonal Polynomials

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

A.1 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 $L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), \tilde{w}(x)\,\mathrm{d}x)$ 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

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

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

with

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

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 $\omega_j = \frac{\tilde{\omega}_j}{\sqrt{\pi}}$.

A.2 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 $L^2([-1,1], \mathscr{B}([-1,1]), \mathrm{d}x)$, 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

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

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

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

so that the germ ξ is defined by

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

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

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

and consequently the Legendre polynomials also form a complete orthogonal system in

$L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathrm{d}P_\xi)$ with

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

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 $\omega_j = \frac{\tilde{\omega}_j}{2}$.

A.3 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 $L^2([-1,1], \mathscr{B}([-1,1]), \tilde{w}(x)\,\mathrm{d}x)$ with weighting function

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

More specifically,

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

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

$$f_\theta(\vartheta;\alpha,\beta) = \frac{1}{B(\alpha,\beta)}\vartheta^{\alpha-1}(1-\vartheta)^{\beta-1}\mathbb{1}_{[0,1]}(\vartheta).^{25}$$

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

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

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

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

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 $\omega_j = \frac{2^{1-\alpha-\beta}}{B(\alpha,\beta)}\tilde{\omega}_j$.

A.4 Generalized Laguerre Polynomials

Generalized Laguerre polynomials are defined by the recurrence relation

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

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

They form a complete orthogonal system on $L^2([0,\infty), \mathscr{B}([0,\infty)), \tilde{w}(x)\,\mathrm{d}x)$ 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 $\theta \sim \text{Gamma}(\alpha,\beta)$, with shape parameter α and rate parameter β is given by

$$f_\theta(\vartheta;\alpha,\beta) := \frac{\beta^{\alpha}}{\Gamma(\alpha)}\vartheta^{\alpha-1}e^{-\beta\vartheta}\mathbb{1}_{[0,\infty)}(\vartheta).^{26}$$

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

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

and that the generalized Laguerre polynomials $\{La_n^{(\alpha-1)}\}_{n\in\mathbb{N}_0}$ also form a complete orthog-

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

onal system in $L^2(\mathbb{R}, \mathscr{B}(\mathbb{R}), \mathrm{d}P_\xi)$ with

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

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 $\omega_j = \frac{\tilde{\omega}_j}{\Gamma(\alpha)}$.

B 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_ξ. Then (13) becomes

$$\hat{y}_\alpha = \|q_\alpha\|_{L^2}^{-2} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} h(\psi(s_1,\dots,s_k)) q_{1\alpha_1}(s_1)\dots q_{k\alpha_k}(s_k) w_1(s_1)\dots w_k(s_k)\,\mathrm{d}s_1\dots\mathrm{d}s_k. \quad (17)$$

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

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

or equivalently taking care of duplicate terms in the difference operators

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

Applying the Smolyak-Gauss-quadrature rule to (17) in particular yields the approximation

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

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