Automated Solution of Heterogeneous Agent Models

David Childers[1*]

November 3, 2019

Abstract

In this paper I present and analyze a new linearization based method for automated solution of heterogeneous agent models with continuously distributed heterogeneity and aggregate shocks. The approach is based on representation of the model equilibrium conditions as a system of smooth functional equations in terms of endogenously time-varying distributions and decision rules. Taking the value of these functions at a set of grid points as arguments, the equilibrium conditions can then be linearized, interpolated with respect to a set of basis functions, and solved through a procedure relying on automatic differentiation and standard discrete time linear rational expectations solution algorithms. While solution approaches based on linearization of discretized or projected models have achieved substantial popularity in recent years, it has been unclear whether such approaches generate solutions which correspond to that of the true infinite dimensional model. I characterize a broad class of models and a set of regularity conditions which ensure that this is indeed the case: the solution algorithm is guaranteed to converge to the first derivative of the true infinite dimensional solution as the discretization is refined.

The key conceptual result leading to these methods is a recognition that a broad variety of heterogeneous agent models can be interpreted as infinite width deep neural networks [Guss, 2017], constructed entirely by iterated composition of pointwise nonlinearities and linear integral operators along a directed acyclic computational graph. On a theoretical level, this formulation ensures commutativity of differentiation and sampling and so permits construction of approximate functional derivatives without performing direct manual calculations in infinite dimensional space. On a practical level, this permits implementation using existing fast and scalable libraries for automatic differentiation on Euclidean space while maintaining the consistency guarantees derived for solutions based on

*Carnegie Mellon University. The author would like to thank Keshav Dogra and Sushant Acharya for extensive discussion, advice, and criticism related to this work, as well as contributions to the code, and seminar participants at Carnegie Mellon and the Federal Reserve Bank of New York and conference participants at the 2019 Barcelona GSE Summer Forum and the Society for Economic Dynamics 2019 Meetings for comments. The author takes full responsibility for all errors, mistakes, or omissions.

derivatives computed directly in infinite dimensional space in Childers [2018].

In addition to providing precise technical conditions under which this method yields accurate representations, I provide examples and guidelines for how to formulate models to ensure that these conditions are satisfied. These conditions are shown to hold in models which possess smooth conditional densities of idiosyncratic state variables as in the class of heterogeneous agent models formalized in Arellano and Bonhomme [2016] augmented with aggregate shocks, subject to a particular choice of representation of the model equations which can be implemented by a change of variables. Convergence rates for the approximation are derived, depending on the classes of functions defining the nodes in the network and the overall network topology for a variety of choices of interpolation method including polynomials, splines, histograms, and wavelets. The procedure is demonstrated numerically by application to a version of the incomplete markets model of Huggett [1993] with continuously distributed idiosyncratic and aggregate income risk.

1 Introduction

Dynamic structural models of heterogeneous individual consumer or firm behavior have by now become a common tool for empirical microeconomics. Incorporating into these models a stochastically changing environment and allowing this environment to respond endogenously to the aggregated behavior permits the integration of these frameworks with those used to study macroeconomic questions. Due to the high dimensionality of the resulting spaces, the integration of these frameworks is computationally challenging, often requiring the use of heuristic approximation approaches or computational methods specially adapted to particular features of the model environment. This work presents a powerful general purpose framework for building and approximately solving dynamic economic models with heterogeneity and endogenous stochastic aggregate variation over time. The approach enjoys strong theoretical guarantees on approximation quality and is highly modular and flexible, permitting fast and scalable model construction and iteration. The construction which makes this feasible is a realization that a wide variety of dynamic economic models can be built, exactly as modern deep neural network architectures are, by repeated composition of "nodes" built from only two types of map: a linear map, approximable by a matrix, and a nonlinear function applied pointwise to its output.

Models with this structure can be represented as a computational graph. From the user side, this allows automated linearization and solution of the model using a combination of standard automatic differentiation libraries and linear rational expectations solution methods. From the analysis side, this allows constructing approximation guarantees for large and complicated models based on low level conditions on the functions used in the individual nodes which ensure the high level conditions in Childers [2018], and so ensure the accuracy

of the solution in terms of the exact infinite dimensional representation. Among other advantages, this allows the application of the function space methods used there to versions of standard models of heterogeneous consumers and firms and extensions thereof.

The class of models and procedures introduced here combine the structure and solution methods for linearized Dynamic Stochastic General Equilibrium (DSGE) models to describe the behavior of aggregates, which may include cross-sectional distributions and other function valued objects, with fully nonlinear projection methods for individual level intertemporal decision problems and their aggregation up to the distribution level: see Fernández-villaverde et al. [2016] for an overview of both perturbation and projection methods. A variety of solution methods for heterogeneous agents models of this type already exist in the literature, many of which are similarly based on combination of function approximation and perturbation methods, including those of Reiter [2009], Winberry [2016], Ahn et al. [2017], Veracierto [2014], Boppart et al. [2017], Auclert et al. [2019], and Chung [2007]. Unlike Childers [2018], each of these approaches is based on linearization of a discretized or projected approximation of a model, rather than the reverse order. This facilitates the use of standard software, but makes analysis of the quality of the approximation difficult. The structure imposed in the approach introduced here ensures commutativity of these operations and so preserves the advantages of both classes of method. As some care must be taken with the model structure in order to ensure this equivalence, implementation of model solution differs in practice from that advocated in previous work, first in the way the equations defining a model must be written, and second, in some additional preprocessing steps performed by the solution algorithm to map the derivatives to the coefficients of a representation with respect to a set of basis functions. This step is automatable but not performed in existing procedures, which either work directly with derivatives of the model equations at a set of points [Reiter, 2009, Ahn et al., 2017], or represent the model equations in terms of sums of basis functions and take derivatives directly with respect to the coefficients [Winberry, 2016], rather than starting with the derivatives of the functions at a set of grid points and then interpolating, as is done here. The presence of guarantees relying on modifications of existing procedures should not be taken to indicate that existing procedures may not also exhibit strong performance, only that the additional structure here facilitates analysis by highlighting a set of features which can ensure algorithm performance.

While highly flexible, the class of model structures covered does impose some economically meaningful restrictions. Most important of these is the existence of continuous conditional densities for agent actions and states, which may require the incorporation of additional sources of stochastic heterogeneity beyond those included in standard dynamic models. Section 2 describes a benchmark class of heterogeneous agent models, nesting the class described in Arellano and Bonhomme [2016], which satisfies these conditions, shows how models of this type can be represented so as to be able to apply the methods introduced here, and provides as an example a version of the model of Huggett [1993] with con-

tinuous heterogeneity. The main section, 3, describes mathematically a much broader class of models defined in terms of directed acyclic graphs of nodes taking functions as inputs and outputs, introduces an algorithm for linearization and model solution, and describes conditions which ensure guarantees on the accuracy of the output of this algorithm. Section 4 demonstrates the application of this method to models of the type described in Section 2, and Section 5 concludes with some discussion of implications of the results for alternative methods. Appendix A discusses in depth how to choose the set of basis functions used with the procedure based on the properties of the model to be solved, and provides lemmas showing how these choices affect the convergence properties of the algorithm in Section 3. Appendix B collects supplementary technical results and Appendix C collects all proofs.

Regarding notation, $\|.\|$ will denote the norm of a Banach space, in the absence of a subscript specifying the space; e.g. $\|f\|_p = (\sum_i f(i)^p)^{\frac{1}{p}}$ or $\|f\|_{L^p} = \left(\int f(x)^p dx\right)^{\frac{1}{p}}$, applied to vectors it will denote Euclidean norm, to functions will denote L^2 norm, and applied to operators will denote operator norm $\|A\|_{op} := \sup_{\|f\|_{L^2}=1} \|A[f]\|_{L^2}$. Discretization parameters will be denoted by K, with subscripts in cases with multiple such parameters, and C will denote a constant independent of parameters K but which may change from line to line. "Big O" order notation will be used throughout; e.g. $\epsilon = O(f(K))$ denotes $\frac{\epsilon}{f(K)} \leq C$ for some C independent of K.

2 Heterogeneous Agent Models - A Framework

While heterogeneous agent models have been applied to a variety of contexts including consumer expenditure, financial, and mobility decisions and firm entry, exit, investment, and pricing,[1] many models in this class retain important structural similarities in terms of the types of equations used, which can be exploited to describe general properties which can be applied more broadly. For the purposes of the solution algorithm, a heterogeneous agent model is defined, exactly as in the literature on perturbation methods for DSGE models following, e.g. Schmitt-Grohé and Uribe [2004], as a system of recursive expectational difference equations differentiable with respect to a set of predetermined and jump variables, to be solved for a recursive solution which takes current predetermined variables and stochastic shocks and produces current jump variables and next period predetermined variables. Unlike in this setup, these variables are allowed to be function valued,[2] so the equilibrium conditions are functional equations, which must be functionally differentiable. I will also impose additional regularity conditions on the structure of these functional equations to ensure that they can be well approximated. The key assumption here will be that all

[1] See, respectively Huggett [1993] and Krusell and Smith, Jr. [1998], Kaplan et al. [2016], Childers [2018], Hopenhayn [1992], Khan and Thomas [2008] and Bachmann et al. [2013], Costain and Nakov [2011], for a far from comprehensive sample.

[2] More precisely, elements of a separable Hilbert space.

equilibrium conditions are composed of repeated composition of only two types of maps; either integration or composition with a smooth function. Although model derivations are now standard in the finite dimensional case, one of the more difficult aspects of using perturbation methods to solve a heterogeneous agents model is deriving the equilibrium condition equations and choosing the function-valued variables and their arguments.

Mathematically, following Schmitt-Grohé and Uribe [2004], Childers [2018] I define a model as a recursive system of equations, which I will call "equilibrium conditions," in predetermined and jump variables in current period $(x, y) \in \mathcal{H}_1 = \mathcal{H}_x \times \mathcal{H}_y$ and in future period $(x', y') \in \mathcal{H}_1$

$$\mathbb{E}F(x, y, x', y', \sigma) = 0$$

where $F : \mathcal{H}_1 \times \mathcal{H}_1 \times \mathbb{R}_+ \to \mathcal{H}_2$ is (functionally) differentiable in each argument.

Let $x = (x_1, x_2) \in \mathcal{H}_x = \mathcal{H}_{x_1} \times \mathcal{H}_{x_2}$ be divided into endogenous component x_1 and exogenous component x_2 such that $x_2' = h_2(x_2) + \sigma z'$ for z' a $\mathcal{H}_{x_2}-$valued random element. A recursive solution is a pair of functions $g(x, \sigma) : \mathcal{H}_x \times \mathbb{R}_+ \to \mathcal{H}_y$ and $h(x, \sigma) : \mathcal{H}_x \times \mathbb{R}_+ \to \mathcal{H}_x$ where $h(x, \sigma) = (h_1(x, \sigma), h_2(x_2))$ and $\eta = [0, I_{\mathcal{H}_{x_2}}] : \mathcal{H}_{x_2} \to \mathcal{H}_x$ such that for all x

$$\mathbb{E}F(x, g(x, \sigma), h(x, \sigma) + \sigma\eta z', g(h(x, \sigma) + \sigma\eta z', \sigma), \sigma) = 0$$

I will introduce a method which can calculate a recursive solution with particular properties, including stationarity, by discretization and linearization. In order to build an efficient method which can approximate such a solution numerically with the flexibility to handle heterogeneous agent models, I will impose additional structure on F, moving from general maps between functions to maps which are built from only two simple building blocks. The function valued variables whose law will be determined by the equilibrium of the model, such as a probability density function of individual level states or a policy rule as a function of those states, will be functions of the form, for some domain $\mathcal{S}_1 \subset \mathbb{R}^{d_1}$

$$g(.) : \mathcal{S}_1 \to \mathbb{R}$$

The equilibrium conditions will be maps, or operators, which take such functions as input, and produce similar functions as output, though possibly in different arguments. A simple example of a map which can perform this is composition with a pointwise nonlinearity (called a Neymitski'i operator in the functional analysis literature [Kesavan, 2004]), defined, for some function $f(s_1, s_2, s) : \mathcal{S}_1 \times \mathcal{S}_2 \times \mathbb{R} \to \mathbb{R}$ as a map which takes as input a function $g(.) : \mathcal{S}_1 \to \mathbb{R}$ and produces a function $\mathcal{S}_1 \times \mathcal{S}_2 \to \mathbb{R}$ of the form

$$F_f[g(.)] = f(s_1, s_2, g(s_1)) \tag{1}$$

Another simple example of a map between functions is a (Fredholm) integral operator, which outputs a weighted average of the input function $g(.) : \mathcal{S}_1 \to \mathbb{R}$,

with the weighting at every point given by a kernel function $k(s_2, s_1) : \mathcal{S}_2 \times \mathcal{S}_1 \to \mathbb{R}$, producing a function $\mathcal{S}_2 \to \mathbb{R}$ of form

$$F_k[g(.)] = \int k(s_2, s_1) g(s_1) ds_1 \tag{2}$$

More complicated maps can be made by repeated composition of maps with these simple forms. The simplest example of such a map is the composition of a nonlinearity and then an integral operator to form a map $F_k[F_f[.]]$ that takes as input a function $g(.) : \mathcal{S}_1 \to \mathbb{R}$ and outputs a function $\mathcal{S}_2 \to \mathbb{R}$ of the form

$$\int k(s_2, s_1) f(s_1, s_2, g(s_1)) ds_1 \tag{3}$$

Such a mapping is referred to in the functional analysis literature as a Hammerstein integral operator [Kesavan, 2004]. In discrete form, maps of this kind are ubiquitous in models of trade and economic geography (see Allen et al. [2015]), and they have also attracted attention in the literature on functional data analysis, where the form is referred to as a functional additive model [McLean et al., 2014], as it nests the standard generalized additive model in the case where s_1 is discrete.

More generally, any structure composed of nested pointwise nonlinearities and integral operators can be used. In what follows, I will in fact impose the restriction that models are built only by repeated composition of maps of the above two forms, (1) and (2). Such structures are extremely expressive. In the finite dimensional case, feedforward deep neural networks (see Goodfellow et al. [2016] for overview) are defined as iterated compositions of linear and pointwise nonlinear maps, and can approximate any continuous function [Hornik et al., 1989]. In the case where inputs are functions, rather than finite dimensional vectors, the corresponding structures, referred to by Guss [2017] as "Deep Function Networks", retain a similar universal approximation property. The structure will in fact allow even greater generality, as rather than restricting to compositions which can be ordered into layers in a tree structure, it suffices to allow compositions to be defined in terms of a general directed acyclic graph (DAG), with nodes defined by integral operators or nonlinearities, and edges passing the output of one node to the input of another. This permits any sequence of compositions which does not result in a cycle, including layered structures but also those which cannot be uniquely ordered, such as "residual" or "skip connections" [He et al., 2016].

A major advantage of using models with the required form is that approximation of and computation with maps in the above form is simple and composes straightforwardly. In place of an input function $g(.) : \mathcal{S}_1 \to \mathbb{R}$, one can use a discretized form, at grid points $\{s_{1k}\}_{k=1}^{K_1}$, replacing the function with a vector $\{g(s_{1k})\}_{k=1}^{K_1}$, and likewise discretize $\mathcal{S}_2$ using points $\{s_{2k}\}_{k=1}^{K_2}$. Using this representation, a pointwise nonlinearity of form (1) produces a map from a K_1 vector to a length $(K_1 \times K_2)$ vector with $k_1 \times k_2$ element

$$f(s_{1k_1}, s_{2k_2}, g(s_{1k_1}))$$

and, using quadrature weights $\{\pi_k^{s_1}\}_{k=1}^{K_1}$ and the same grid points, an integral operator of form (2) produces a map from $K_1 \times K_2$ vectors $g(.,.)$ on $\{s_{1k}\}_{k=1}^{K_1} \otimes \{s_{2k}\}_{k=1}^{K_2}$ to length K_2 vectors with j^{th} element

$$\sum_{k=1}^{K_1} \pi_k^{s_1} k(s_{2j}, s_{1k}) g(s_{1k}, s_{2k})$$

Repeatedly composing these two forms, which always produce vectors on a grid which is fixed ex ante, produces a network mapping vectors to vectors, nesting traditional finite dimensional neural networks. To produce a linearized model, the network can be differentiated through by the chain rule as implemented in standard automatic differentiation software. In principle, one could take the resulting Jacobian matrices and apply linear rational expectation solution algorithm to produce an approximate model solution. In practice, additional post-processing steps will be introduced, based on interpolation of grid points to full functions, which can be used to allow high accuracy with a substantially coarser discretization. Under a set of conditions on the structure of the network and the functions of which it is comprised, I will show that such a procedure can have highly desirable accuracy properties.

While networks of the described form may approximate nearly arbitrary continuous maps (see Guss [2017], Guss and Salakhutdinov [2019] for precise universality claims in weak and uniform norm topologies, respectively), it is natural to consider whether models of interest naturally take such a form, or if they must be modified in order to use tools based on the structure. Fortunately, the structure nests many standard economic models exactly, though formatting a model in order to fit the form may require some transformation of variables. Many forms which are ruled out by the limitation to maps of classes (1) and (2) are excluded deliberately, as they often result in models which are difficult or impossible to differentiate or approximate well by a finite set of basis functions. The most common such form is the evaluation (or Dirac delta) operator: given a function $f(.) : \mathcal{S}_2 \to \mathcal{S}_1$, it takes an input function $g(s_1)$ and produces a function $\mathcal{S}_2 \to \mathbb{R}$ of form

$$\delta_f[g(.)] = g(f(s_2)) \tag{4}$$

Such evaluation operators typically come up in intertemporal decision problems, where a value or policy tomorrow must be evaluated at a point determined by a decision today. While superficially similar to a pointwise nonlinearity, the reversal of the order of composition is quite troublesome. On a practical level, it requires evaluating an input function at points other than a predefined set of grid points.[3] At the level of performance guarantees, this creates issues with approximation accuracy because some functions $g(.)$ can be very sensitive to

[3]This issue is sidestepped in existing perturbation and projection approaches by using a representation of the function $g(s_1)$ defined off of a grid either by numerical interpolation, as in Reiter [2009], or using basis functions, as in Winberry [2016]. This makes the form implementable, but theoretical issues remain. It is still an open problem to describe circumstances under which such an approach can ensure consistent approximation.

small shifts in their input. For example, if they have a large spike at one point, small movements in the argument can lead to large changes in the output. Moreover, because the functions $g(.)$ are determined endogenously in equilibrium, it is hard to ensure that they will not be so poorly behaved. At a technical level, this sensitivity reflects the fact that the evaluation operator is not continuous in $g(.)$ unless $g(.)$ is restricted to a small class, and this discontinuity can confer large sensitivity to approximation errors. The precise concern is that maps of this form are not *compact*, and so cannot be consistently approximated in operator norm by any linear map which maps a finite dimensional basis function representation to a finite dimensional basis function representation. This compactness condition was key to the solution accuracy guarantees of Childers [2018] for a projection approximation of the solution to models of this type. While problematic when occuring alone, in many models, maps of this form occur only when composed with other (compact) operators, and so compactness can be restored by applying a change of variables; in others, one can change the model by adding additional sources of noise and then apply a change of variables. This technique will be applied in subsequent sections to construct classes of models which do retain compactness and so can be approximated by projection methods.

Beyond the restrictions imposed by building models only from components of forms (1) and (2), ensuring consistent approximation will impose some restrictions on how they are composed. The point of such restrictions is to ensure that the linearized form of the model is made up only of maps which can be well approximated by function interpolation. This is easily achieved if the linearized model is composed of kernel integral operators (2), for which uniformly accurate approximation of their action on the space of functions is implied by uniform approximation of the kernel function $k(s_2, s_1)$, which can be guaranteed by standard interpolation methods. Applying the chain rule, any directed path composed of maps of types (1) and (2) containing at least one map of type (2) will have a derivative in this form. The identity map is also easily approximated by replacing it with an identity matrix. The only concern in a model of this type is the presence of chains along the graph containing only nonlinear maps of type (1). These result in a derivative which takes the form of a multiplication operator. Given a function $f(.) : \mathcal{S}_1 \to \mathbb{R}$, it takes an input function $g(.) : \mathcal{S}_1 \to \mathbb{R}$ and produces a function $\mathcal{S}_1 \to \mathbb{R}$ of form

$$F_f[g(.)] = f(s_1) \cdot g(s_1) \tag{5}$$

The reason such a map is not innocuous is that when the input is represented by a finite representation, the resulting output need not be well approximated in the same space. Consider the multiplication map with $f(s_1) = s_1$. If $g(s_1)$ is approximated by an order K polynomial, $s_1 g(s_1)$ must be a polynomial of order $K + 1$, and this is true for any K. As a result, a square matrix of any size will not (uniformly) accurately represent the behavior of this map.[4]

[4]Non-square matrices could be used instead; see Adcock et al. [2014] for discussion of approximation by rectangular matrices and Childers [2016, Ch. 3] for a solution approach

However, if the function $f(s_1)$ is nonzero, one can often get rid of maps of this form by dividing the equation through by $f(s_1)$, so that the multiplication becomes an identity and any kernel integral maps remain so, but now with the kernel function divided by $f(s_1)$. When feasible, this step can be checked and performed automatically. One may also be able to avoid these cases by changing choice of functions used as state variables, eg, if $g(s_1)$ appears inside $f(g(s_1))$, one can define $\ell(s_1) := f(g(s_1))$ directly as the state variable. Examples of this approach are also provided. Precise sufficient conditions ensuring that a model takes a form which can be represented in the restricted form are provided, and verified in realistic examples, in the subsequent sections

In order to see how this abstract setting nests practically relevant heterogeneous agent models, and how to set them up so that the methods introduced here can be used, I first introduce a specialized setup with greater economic structure.

2.1 Canonical Class of Heterogeneous Agent Models

In what follows, I describe a class of heterogeneous agent models, corresponding to those studied in Arellano and Bonhomme [2016], Arellano et al. [2017], in which one can systematically derive a choice of functions and arguments sufficient to fully characterize the model, classify them as predetermined or contemporaneous, and derive a formula for the law of motion of the function-valued states which is sufficiently regular.

Ensuring that a model is sufficiently regular may often require the inclusion of additional sources of idiosyncratic heterogeneity beyond those traditionally included in non-quantitative heterogeneous agents models, in order to ensure the existence of non-singular conditional densities. However, given these conditions, one has the advantage that the nonparametric identification arguments of Arellano and Bonhomme [2016] may then easily be verified (given appropriate data) and so the version of their estimation procedure permitting time-varying functions may be used to recover period-by-period estimates of the function-valued state variables of interest.

In what follows, I adopt the notation of Arellano et al. [2017]. For notational convenience I omit time-invariant unobserved heterogeneity (i.e., "fixed effects"), which may trivially be added as an argument to all the functions without changing any of the subsequent results. At the individual level, I introduce the variables $(Y_{it}, X_{it}) \in (\mathcal{X}, \mathcal{Y}) \subseteq \mathbb{R}^{n_y+n_x}$ which are the individual level "choice" and "predetermined" variables, respectively, not to be confused with the same designations for aggregate states. For example, in a consumption-savings model, Y_{it} might contain consumption of agent i in time t, as well as decisions such as labor supply or spending on durable goods, while X_{it} might contain asset holdings and (persistent and transitory components of) productivity or labor income. In addition to these variables, I consider vectors of auxiliary variables (U_{it}, ϵ_{it}) which

based on this method. This introduces a variety of complications which are avoided by the methods introduced here.

are independent of each other and across time and individuals, with absolutely continuous densities f_U and f_ϵ, respectively and whose functional role is to introduce stochastic variation and ensure existence of conditional densities of X_{it} and Y_{it}, respectively. I require that $(U_{it}, \epsilon_{it}) \in (\mathcal{U}, \epsilon) \subseteq \mathbb{R}^{n_u, n_\epsilon}$ with $n_u \geq n_x$. An example of such shocks might be, for U_{it}, stochastic income or asset returns which make next period asset holdings stochastic conditional on decisions today, and for ϵ_{it}, idiosyncratic shocks to preferences or constraints which cause consumers with the same current assets to engage in different spending and savings behavior.

For the purpose of solving the model via the proposed algorithms, the variable U_{it} is essential while ϵ_{it} can be omitted entirely if not economically relevant. While both are necessary for the estimation procedure and identification arguments of Arellano and Bonhomme [2016], which require existence of a conditional density for both X_{it} and Y_{it}, the procedure only needs existence of a nonsingular conditional density for the former. Singularity of the latter density is not a computational problem, and in fact may have desirable computational properties as omitting ϵ_{it} would reduce the dimensionality of the endogenous functions. Such a simplification cannot be made with respect to the X_{it} variables. This nonsingularity condition rules out models where idiosyncratic states evolve deterministically over time. In some cases this can be avoided by a redefinition of variables; in others, it requires the addition of novel idiosyncratic shocks.

At the aggregate level, I introduce a set of aggregate variables $P_t = (P_{1t}, P_{2t}) \in \mathcal{H}_{P_1} \times \mathcal{H}_{P_2}$, which represent the endogenous and exogenous aggregate variables, respectively, which do not depend on i, may be of arbitrary dimension, and may even be functions on $\mathcal{X}$, $\mathcal{Y}$, or ϵ to be evaluated at X_{it}, Y_{it}, or ϵ_{it}. Examples of P_{1t} might include market prices and aggregate quantities or other endogenous features of the environment determined collectively by the distribution of agent behavior. P_{2t} may include aggregate shocks which may affect the preferences, technologies, or other aspects of the environment facing many agents. The procedure will treat arguments with *it* subscripts (idiosyncratic variables) differently from those with only t subscripts (aggregate variables); projection will be applied to functions of idiosyncratic variables, while perturbation will be applied to aggregate variables.

The predetermined state variables X_{it} are assumed to follow a law of motion, given exogenously by the model itself.

$$X_{it+1} = Q(Y_{it}, X_{it}, P_t, P_{t+1}, U_{it+1})$$

where Q is a known, given function, where Y_{it}, X_{it} and P_t (but not U_{it+1} or P_{t+1}) are presumed to be known at time t. An example of such a function would be the law of motion for an exogenous idiosyncratic productivity or wage process, or an asset accumulation equation determining next period asset holdings based on current saving decisions. Note that both time t and time $t+1$ aggregate variables may be included in this rule. For example, the value of an agent's time $t+1$ asset holdings might depend on time $t+1$ asset prices.

In contrast, the choice variables Y_{it} are determined contemporaneously by a function of current period predetermined variables

$$Y_{it} = g_t(X_{it}, \epsilon_{it})$$

which in general is not known directly but must be determined endogenously as, for example, the solution to a set of optimization conditions, and may vary over time. An example would be a consumption or labor supply decision rule as a function of current income and/or assets.

Given such a model, I can characterize the set of function valued states needed to describe the model and their laws of motion.

As indicated by the time subscript, the function $g_t(.,.)$ can be used as a function-valued state variable in the aggregate model, living on the space of functions over the joint support of X_{it} and ϵ_{it}. Its law of motion may be written implicitly as the solution of a set of optimization conditions (either through an Euler equation or a Bellman equation). This function will be a jump variable with respect to the aggregate variation. In many cases, it is sensible to use some other function of arguments X_{it}, ϵ_{it} instead of the decision rule as the function-valued state variable, such as a value function (as used in the economic geography model in Childers [2018]) or the "parameterized expectations" function [Christiano and Fisher, 2000], so long as the decision rule determining Y_{it} can be written as a known function of the real-valued output of this function. This can be used to handle models with occasionally binding constraints on individual-level decisions while maintaining smoothness of the function-valued state.

There are a variety of ways to derive such a decision rule, but the most common is from an intertemporal optimization problem. A version of such a problem consistent with this setup is as follows. An agent solves Bellman equation

$$V_t(X_{it}, \epsilon_{it}) = \max_{Y_{it} \in S(X_{it}, P_t, \epsilon_{it})} u(Y_{it}, X_{it}, P_t, \epsilon_{it}) + \beta \mathbb{E}_t \int \int V_{t+1}(Q(Y_{it}, X_{it}, P_t, P_{t+1}, U_{it+1}), \epsilon_{it+1}) \cdot$$

$$f_U(U_{it+1}) f_\epsilon(\epsilon_{it+1}, P_{t+1}) dU_{it+1} d\epsilon_{it+1}$$

where $u(.,.,.,.)$ describes contemporaneous utility and $S(.)$ contemporaneous constraints, and $\mathbb{E}_t$ describes the conditional expectation over macroeconomic aggregate P_{t+1}. The additively separable discounted form of the decision problem is meant to be illustrative; nonseparable preferences may be accommodated at the cost of a minor increase in notational complexity. The variable ϵ_{it}, representing contemporaneous shocks to preferences or constraints, may often be solved out by use of the "ex ante" value function, defined as the expectation of $V(X_{it}, \epsilon_{it})$ with respect to the distribution of ϵ_{it}, as is common in the case of dynamic discrete choice problems following Rust [1987]. It may also be omitted entirely, which causes no computational issues, though may lead to stochastic singularity and its associated econometric difficulties. Note that P_t is not denoted as entering directly as an argument to the value function. This is not

because the value function does not depend on it; instead, it is left implicit in the t subscript. Less innocuously, aggregate variables also do not enter into the density f_U. This is important for the computations but not in any way economically restrictive as any effect of aggregates on the distribution of this idiosyncratic shock can be expressed through the functional form of Q.

Assuming differentiability and concavity of the utility function and an interior solution,an optimal decision rule $Y_{it} = g_t(X_{it}, \epsilon_{it})$ satisfies equations

$$V_t(X_{it}, \epsilon_{it}) = u(g_t(X_{it}, \epsilon_{it}), X_{it}, P_t, \epsilon_{it}) +$$

$$\beta \mathbb{E}_t \int \int V_{t+1}(Q(g_t(X_{it}, \epsilon_{it}), X_{it}, P_t, P_{t+1}, U_{it+1}), \epsilon_{it+1}) \cdot$$

$$f_U(U_{it+1}) f_\epsilon(\epsilon_{it+1}, P_{t+1}) dU_{it+1} d\epsilon_{it+1}$$

and

$$u_Y(g_t(X_{it}, \epsilon_{it}), X_{it}, P_t, \epsilon_{it}) +$$

$$\beta \mathbb{E}_t \int \int \frac{\partial}{dX_{it+1}} V_{t+1}(Q(g_t(X_{it}, \epsilon_{it}), X_{it}, P_t, P_{t+1}, U_{it+1}), \epsilon_{it+1}) \cdot$$

$$\frac{\partial}{dY_{it}} Q(g_t(X_{it}, \epsilon_{it}), X_{it}, P_t, P_{t+1}, U_{it}) f_U(U_{it+1}) f_\epsilon(\epsilon_{it+1}, P_{t+1}) dU_{it+1} d\epsilon_{it+1} = 0$$

In order to be consistent with the conditions of the method, the above conditions must be transformed to be used directly in the method to be described. Supposing I let the functions $g_t(X_{it}, \epsilon_{it})$ and $V_t(X_{it}, \epsilon_{it})$ be function valued variables, the above equations require evaluating V_t at the value of another function and using its derivative, both of which correspond to application of operations other than integration and composition with a function; from a technical perspective, these transformations are non-compact and so are not approximable directly. Combining these two equations into a single Euler equation removes the value function as a variable, but generally is defined in terms of $g_{t+1}(Q(g_t(X_{it}, \epsilon_{it}), X_{it}, P_t, P_{t+1}, U_{it+1}), \epsilon_{it+1})$, which again involves a noncompact evaluation operator.

The solution to this is simple, so long as there exists a shock term U_{it+1} with smooth density which enters the transition rule of X_{it}. To remove the nested evaluations, one can apply a change of variables so that V_t (respectively g_{t+1}) is instead integrated over the smooth conditional density of X_{it+1}. To remove the differentiation, one can apply integration by parts or, equivalently, first perform the change of variables and then take the first order condition with respect to X_{it+1}. Applying the transform $X_{it+1} = Q(g_t(X_{it}, \epsilon_{it}), X_{it}, P_t, P_{t+1}, U_{it+1})$ with inverse $U_{it+1} = Q^{-1}_{g_t(X_{it}, \epsilon_{it}), X}(X_{it+1})$, which has $P_t,\ P_{t+1}$ arguments omitted for brevity, this results in equivalent equations

$$V_t(X_{it},\epsilon_{it}) = u(g_t(X_{it},\epsilon_{it}),X_{it},P_t,\epsilon_{it}) + \beta\mathbb{E}_t\int\int V_{t+1}(X_{it+1},\epsilon_{it+1})\cdot \quad (6)$$
$$f_U(Q^{-1}_{g_t(X_{it},\epsilon_{it}),X}(X_{it+1}))f_\epsilon(\epsilon_{it+1},P_{t+1})\cdot$$
$$\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g_t(X_{it},\epsilon_{it}),X}(X_{it+1})\right| dX_{it+1}d\epsilon_{it+1}$$

and

$$u_Y(g_t(X_{it},\epsilon_{it}),X_{it},P_t,\epsilon_{it}) \quad (7)$$
$$+\beta\mathbb{E}_t\int\int V_{t+1}(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU_{it+1}}f_U(Q^{-1}_{g_t(X_{it},\epsilon_{it}),X}(X_{it+1}))\cdot$$
$$\frac{\partial}{dY}Q^{-1}_{g_t(X_{it},\epsilon_{it}),X}(X_{it+1})f_\epsilon(\epsilon_{it+1},P_{t+1})\cdot$$
$$\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g_t(X_{it},\epsilon_{it}),X}(X_{it+1})\right| dX_{it+1}d\epsilon_{it+1} = 0$$

It is this latter system of equations which is consistent with the conditions of the baseline perturbation approach so long as the associated components are appropriately regular. In cases in which the constraint $Y_{it} \in S(X_{it},P_t,\epsilon_{it})$ is occasionally binding, a Lagrange multiplier may need to be added as a state, along with a complementary slackness condition as an equation. In these cases, as well as in others, it may be possible to combine the set of equations into a single Euler equation, with only the policy rule $g_t(X_{it},\epsilon_{it})$ as a state variable, eliminating the value function and, if included, the Lagrange multiplier from the system. See the example version of the consumption savings model with borrowing constraint in section (4) for an example of this process. Alternately, in those cases where the first order conditions can be solved explicitly for a solution in terms of the value function, the policy rule $g_t(X_{it},\epsilon_{it})$ can be eliminated as a state variable, and replaced where it appears in subsequent calculations by a composition of functions.

To describe how individual behavior aggregates, I take the cross sectional distribution over individuals of X_{it}, represented by its density, denoted $f_{X,t}(.)$ as a state variable. In order to assure that such a density exists and has an easily representable law of motion, I consider first the case where $n_x = n_u$, and assume that for any fixed $(Y_{it},X_{it}) = (y,x)$, the function $Q_{y,x}(U_{it+1}) := Q(y,x,P_t,P_{t+1},U_{it+1})$ is invertible in U_{it+1} with inverse $Q^{-1}_{y,x}(X_{it+1})$ which is differentiable in argument X_{it+1}. Given a conditional density $f(X_{it+1}|Y_{it},X_{it})$, the unconditional density of X_{it+1} can be derived from its previous value by first substituting in the law of Y_{it} to obtain

$$f(X_{it+1}|g_t(X_{it},\epsilon_{it}),X_{it}) = f(X_{it+1}|X_{it},\epsilon_{it})$$

which can be integrated over the conditioning set to obtain the unconditional density as

$$f_{X,t+1}(X_{it+1}) = \int\int f(X_{it+1}|X_{it},\epsilon_{it})f_{X,t}(X_{it})f_\epsilon(\epsilon_{it},P_t)dX_{it}d\epsilon_{it}$$

The conditional density $f(X_{it+1}|Y_{it}, X_{it})$ can be derived from the law of motion for X_{it} by integrating out the shock and performing a change of variables to ensure that this object can be represented as a smooth density:

$$\begin{aligned} f(X_{it+1}|Y_{it}, X_{it}) &= \int f(X_{it+1}|Y_{it}, X_{it}, U_{it+1}) f_U(U_{it+1}) dU_{it+1} \\ &= \int \delta(X_{it+1} - Q(Y_{it}, X_{it}, U_{it+1})) f_U(U_{it+1}) dU_{it+1} \\ &= f_U(Q^{-1}_{Y_{it}, X_{it}}(X_{it+1})) \left| \det \frac{\partial}{dX_{it+1}} Q^{-1}_{Y_{it}, X_{it}}(X_{it+1}) \right| \end{aligned}$$

where $\delta(.)$ is the Dirac delta distribution. Substituting in the above, I obtain the law of motion for $f_{X,t}$ as

$$\begin{aligned} f_{X,t+1}(X_{it+1}) = \int \int f_U(Q^{-1}_{g_t(X_{it}, \epsilon_{it}), X_{it}}(X_{it+1})) \cdot \\ \left| \det \frac{\partial}{dX_{it+1}} Q^{-1}_{g_t(X_{it}, \epsilon_{it}), X_{it}}(X_{it+1}) \right| f_{X,t}(X_{it}) f_\epsilon(\epsilon_{it}, P_t) dX_{it} d\epsilon_{it} \end{aligned} \quad (8)$$

In the case where $n_u > n_x$, invertibility can be replaced with surjectivity: so long as one can find a map $Q^{-1}_{y,x}(X_{it+1})$ injective into the support of U_{it+1} which is differentiable, one can perform a change of variables over the range of $Q^{-1}_{x,y}$, leaving the complement of this support unchanged and proceed as below. For example, if $U_{it+1} = (U^1_{it+1}, U^2_{it+1})$ with density $f_U(U^1_{it+1}, U^2_{it+1})$ and $Q(y, x, U^1_{it+1}, U^2_{it+1})$ is invertible in U^1_{it+1}, one can use $Q^{-1}_{Y_{it}, X_{it}}(X_{it+1}, U^2_{it+1})$ and the law of motion becomes

$$\begin{aligned} f_{X,t+1}(X_{it+1}) = \int \int \int f_U(Q^{-1}_{g_t(X_{it}, \epsilon_{it}), X_{it}}(X_{it+1}, U^2_{it+1}), U^2_{it+1}) \cdot \\ \left| \det \frac{\partial}{dX_{it+1}} Q^{-1}_{g_t(X_{it}, \epsilon_{it}), X_{it}}(X_{it+1}, U^2_{it+1}) \right| f_{X,t}(X_{it}) f_\epsilon(\epsilon_{it}, P_t) dX_{it} d\epsilon_{it} dU^2_{it+1} \end{aligned} \quad (9)$$

If surjectivity holds but U_{it+1} does not have such an a priori decomposition, a similar formula may be derived, albeit with a more complicated change of variables formula.

In the special case where $Q(.)$ is additive in U_{it+1}, $X_{it+1} = Q(Y_{it}, X_{it}, P_t, P_{t+1}) + U_{it+1}$ the determinant term is just a constant, and the inverse is given as $U_{it+1} = X_{it+1} - Q(Y_{it}, X_{it}, P_t, P_{t+1})$, so the transition equation is just

$$f_{X,t+1}(X_{it+1}) = \int \int f_U(X_{it+1} - Q(g_t(X_{it}, \epsilon_{it}), X_{it}, P_t, P_{t+1})) f_{X,t}(X_{it}) f_\epsilon(\epsilon_{it}, P_t) dX_{it} d\epsilon_{it}$$

In many cases, one can save on arguments by creative definition of what X_{it} is, pushing other variables into U, which doesn't enter as an argument to the function-valued states. This may be model-specific, but should be exploited when possible.

Beyond the equations determining the decision rule and the implied law of motion for the cross sectional distribution, there may be additional equations describing how aggregate variables are determined endogenously from individual cross section. These can take a variety of forms, including market clearing constraints, aggregate production functions, government policy responses, or much more complicated systems such as a complete intratemporal trade model, including, for example, any of the CES class of trade models described in Arkolakis et al. [2012] or economic geography models as in Krugman [1991]. For the purposes of the methods introduced here, any such system can be represented as an equation of the form

$$F(f_{X,t}(X_{it}), g_t(X_{it}, \epsilon_{it}), P_t) = 0 \tag{10}$$

where F is defined in terms of integration and composition of its arguments. For example, if Y_{it} is quantity demanded by consumer i of a good or service, and supply is fixed at 0, the equation $F(f_{X,t}(X_{it}), g_t(X_{it}, \epsilon_{it}), P_t) := \int\int g_t(X_{it}, \epsilon_{it}) f_{X,t}(X_{it}) f_\epsilon(\epsilon_{it}, P_t) dX_{it} d\epsilon_{it} = 0$ could represent market clearing in this market.

Finally, exogenous sources of aggregate variation which I denote Z_t, enter the model directly only through P_{2t}, which follows a law of motion of the form

$$P_{2t+1} = h_P(P_{2t}, \sigma Z_{t+1}) \tag{11}$$

where h_P is an exogenously given function and σ is a scalar parameter indexing the magnitude of the shocks.

One important concern when introducing aggregate shocks into a model is that the timing of their realization can affect whether a variable should be treated as predetermined or a jump variable. As long as $Q(Y_{it}, X_{it}, P_t, P_{t+1}, U_{it+1})$ is a non-constant function of P_{t+1}, then the distribution $f_{X,t+1}(X_{t+1})$ will in part be determined by these same-period variables. As a result, $f_{X,t+1}(X_{t+1})$ must be considered a jump variable rather than a predetermined variable. A simple way to handle this, if it occurs, is to add additional lagged versions Lf_X and Lg of the cross section distribution and policy function to the model, which are predetermined variables, add the equations

$$\begin{aligned} Lf_{X,t+1}(X_{it}) &= f_{X,t}(X_{it}) & (12)\\ Lg_{t+1}(X_{it}, \epsilon_{it}) &= g_t(X_{it}, \epsilon_{it}) & (13)\\ LP_{t+1} &= P_t & (14) \end{aligned}$$

and replace the Kolmogorov Forward Equation (8) with

$$\begin{aligned} f_{X,t}(X_{it}) = \int\int f_U(Q^{-1}_{Lg_t(X_{it-1},\epsilon_{it-1}),X_{it-1}}(X_{it}))\cdot \\ \left|\det \frac{\partial}{dX_{it}} Q^{-1}_{Lg_t(X_{it-1},\epsilon_{it-1}),X_{it-1}}(X_{it})\right| Lf_{X,t}(X_{it-1}) f_\epsilon(\epsilon_{it-1}, LP_t) dX_{it-1} d\epsilon_{it-1} \end{aligned} \tag{15}$$

or if this rule is defined by Equation (9), perform the analogous replacement.

Together, equations (6), (7), (10), (11), (13), and (15) fully define a heterogeneous agents model. The predetermined state variables are $Lf_{X,t}(X_{it})$, $Lg_t(X_{it}, \epsilon_{it})$, LP_t, and P_{2t} and jump variables are $f_{X,t}(X_{it})$, $g_t(X_{it}, \epsilon_{it})$, P_{1t}, and $V_t(X_{it}, \epsilon_{it})$.

2.2 Example: A Version of Huggett [1993]

To see how the above framework allows expression of models of interest, including the consumption savings framework which has been dubbed the "standard incomplete markets model" for its use as a workhorse in studying consumer behavior [Heathcote et al., 2009], I express a consumer model with borrowing constraint and one bond as in Huggett [1993] in terms of the above framework. In order to represent the individual states in terms of continuous functions, I assume individual income is drawn from a distribution with absolutely continuous density. To illustrate the role and timing of aggregate shocks, I incorporate both a shock to aggregate income, shared by all individuals, and an "anticipated uncertainty shock", which leads to an anticipated increase in the next period dispersion of idiosyncratic income.

A complete description of the model is as follows. Households solve

$$\max E_0 \sum_{t=0}^{\infty} \beta^t \frac{(c_{it})^{1-\gamma}}{1-\gamma} \tag{16}$$

$$\text{s.t. } w_{it+1} = R_t(w_{it} + z_t - c_{it}) + y_{it+1} \tag{17}$$

$$R_t(w_{it} - c_{it}) \geq \underline{a} \tag{18}$$

where $\underline{a} < 0$ and

$$z_{t+1} = \rho_z z_t + \epsilon^z_{t+1} \tag{19}$$

and $y_{it+1} \sim \tilde{g}(\cdot|\sigma_t)$ where

$$\ln \sigma_{t+1} = \rho_\sigma \ln \sigma_t + \epsilon^\sigma_{t+1} \tag{20}$$

Markets clear, with the bond in 0 net supply.

$$\int (w_{it} + z_t - c_{it}) di = 0 \tag{21}$$

In this model, the individual predetermined variable X_{it} is w_{it}, cash on hand, with density $f_{X,t}(.)$ denoted as $m_t(w_{it})$, and the individual decision variable Y_{it} is c_{it}, spending. Note that the use of cash on hand rather than income as the individual state is deliberate; while income is determined exogenously, cash on hand is a persistent state which evolves stochastically over time, and so will have a continuous conditional density, due to the continuous density of y_{it+1}. In this

model there is no idiosyncratic shock ϵ_{it} to the choice variable.[5] The exogenous aggregates P_{2t} are z_t, σ_t, with shocks $Z_{t+1} = (\epsilon^z_{t+1}, \epsilon^\sigma_{t+1})$ and h_P is given by equations (19) and (20). The endogenous aggregate P_{1t} is R_t, the interest rate.

In order to represent the model in the above form the idiosyncratic shock to the predetermined state must be represented in a nonstandard way. The reason for this is that to be able to invert the transition rule and obtain a map which is a differentiable integral operator in the function valued variables, the density of U_{it} must be time invariant. Although y_{it}, the individual level component of income, represents a shock to X_{it}, due to the fact that aggregate variable affects the density, it cannot be placed in this role. Fortunately, this is resolvable by a simple reparameterization trick which leaves the model itself unchanged. Let U_{it} instead be a variable with time invariant density $q(.)$, and y_{it}, be a function of this and the aggregate variable P (in this case, σ_t). So long as this mapping is invertible, it can be introduced into the transform Q, which is allowed to take aggregate variables as arguments, without loss of generality. There are many acceptable ways to perform this reparameterization, so long as $q(.)$ is time invariant and the combination ensures that $y_{it+1} \sim g(.|\sigma_t)$; here I introduce two.

The first is the inverse CDF transform. Let $U_{it+1} \sim$ Uniform$[0,1]$ and $G^{-1}(.|\sigma_t)$ be the inverse CDF (quantile function) associated to PDF $g(.|\sigma_t)$. Then $y_{it+1} = G^{-1}(U_{it+1}|\sigma_t) \sim g(.|\sigma_t)$, and U_{it+1} can be recovered by applying the CDF transform to y_{it+1}. This transformation can be applied any time a unique quantile transformation exists; if desired, U_{it+1} can be allowed to take any density by including the associated CDF in the transformation. An alternate mapping, which can be used any time the density $g(.|\sigma_t)$ is smooth and is slightly simpler to implement, albeit more difficult to derive, is to define $y_{it+1} = U_{it+1} + s_{it+1}$, where $U_{it+1} \sim q(.)$ and $s_{it+1} \sim g(.|\sigma_t)$ is an additional variable drawn independently. The resulting sum has density equal to the convolution of the two variables, $\tilde{g}(y_{it+1}; \sigma_t) := \int g(s_{it+1}; \sigma_t) q(y_{it+1} - s_{it+1}) dy_{it+1}$, which by appropriate choice of g and q can be made to represent an arbitrary smooth density. Strictly, s_{it+1} should be interpreted as the inverse CDF transform of an additional element of U^2_{it+1}, through which one can apply the multidimensional inverse transform from equation (9), after which a reverse change of variables can be used to recover the density g. Solving for the inverse gives a Bellman equation of form

$$\begin{aligned} V_t(w_{it}) \quad = \quad & \max_{\{c_{it}:\ R_t(w_{it}-c_{it}) \geq \underline{a}\}} \frac{(c_{it})^{1-\gamma}}{1-\gamma} + \mathbb{E}_t \beta \int \int V_{t+1}(w_{it+1}) \cdot \\ & q(w_{it+1} - R_t(w_{it} + z_t - c_{it}) - s_{it+1}) g(s_{it+1}; \sigma_t) ds_{it+1} dw_{it+1} \end{aligned}$$

Due to the presence of a borrowing constraint, a first order condition alone

[5] Such a shock could easily be added by adding an idiosyncratic preference shifter to the utility function or discount rate, reflecting idiosyncratic heterogeneity in demand, liquidity, or patience, or into the borrowing constraint to reflect idiosyncratic variation in access to credit. While such variation would be necessary to match data on individual spending, it is not needed for the computational procedure proposed and adds a dimension to the value function and so is omitted for simplicity of exposition and computation.

is insufficient to characterize an optimal consumption policy rule corresponding to the above Bellman equation; at minimum, one also needs a complementary slackness condition. Following Christiano and Fisher [2000], these optimality conditions can be combined into a single Parameterized Expectations representation of an Euler equation, introducing a variable $\ell_t(w_{it})$ defined as βR_t times the expected marginal utility tomorrow, and then defining the policy rule for consumption as the function which ensures that the constraint is always satisfied, $c(w, \ell, R) := \min\left\{\ell^{-1/\gamma}, w - \frac{\underline{a}}{R}\right\}$. Following those steps results in Euler equation

$$\begin{aligned}\ell_t(w_{it}) = \quad & (22)\\ & \mathbb{E}_t \beta R_t \int\int q(w_{it+1} - R_t(w_{it} + z_t - c(w_{it} + z_t, \ell_t(w_{it}), R_t)) - s_{it+1}) \cdot \\ & g(s_{it+1}; \sigma_t) c(w_{it+1} + z_{t+1}, \ell_{t+1}(w_{it+1}), R_{t+1})^{-\gamma} ds_{it+1} dw_{it+1}\end{aligned}$$

This construction removes the max operator over a set of functions, which is not differentiable, and also replaces all inequalities with strict equalities, allowing the model to be represented solely as a system of equations. It does introduce a min operator in the definition of the consumption rule, but this is simply a pointwise nonlinear function composed with its argument, resulting in a form which is precisely that which can be handled by the proposed method.

Combing these steps to construct the state transition rule for cash on hand, obtain

$$\begin{aligned}m_{t+1}(w_{it+1}) = \quad & (23)\\ & \int\int q(w_{it+1} - R_t(w_{it} + z_t - c(w_{it} + z_t, \ell_t(w_{it}), R_t)) - s_{it+1}) \cdot \\ & g(s_{it+1}; \sigma_t) m_t(w_{it}) ds_{it+1} dw_{it}\end{aligned}$$

which also results in the market clearing condition

$$0 = \int (w_{it} + z_t - c(w_{it} + z_t, \ell_t(w_{it}), R_t)) m_t(w_{it}) dw_{it} \tag{24}$$

Finally, in order to maintain the timing distinctions between predetermined and jump variables, lagged versions of aggregate variables are added of forms Lm_t, $L\ell_t$, LR_t, Lz_t, $L\sigma_t$ defined as

$$\begin{aligned} Lm_{t+1}(w_{it}) &= m_t(w_{it}) \\ L\ell_{t+1}(w_{it}) &= \ell_t(w_{it}) \\ LR_{t+1} &= R_t \\ Lz_{t+1} &= z_t \\ L\sigma_{t+1} &= \sigma_t \end{aligned} \tag{25}$$

and resulting in modified state transition equation (23)

$$m_t(w_{it}) = \int\int q(w_{it} - LR_t(w_{it-1} + Lz_t - c(w_{it-1} + z_t, L\ell_t(w_{it}), LR_t)) - s_{it}) \cdot g(s_{it}; L\sigma_t) Lm_t(w) ds_{it} dw_{it-1} \tag{26}$$

Together, equations (19), (20), (24), (22), (25), and (26) define a system of equations in terms of predetermined variables $Lm_t(w), L\ell_t(w), LR_t, Lz_t, L\sigma_t, z_t, \sigma_t$ and jump variables $m_t(w), \ell_t(w), R_t$. All equations defined are constructed in terms of pointwise nonlinear composition and application of integral operators, and so the algorithm which will be defined in the next sections applies directly to this version of the model.

3 Model Class and Algorithms

In order to construct a solution of a model with a mix of function and scalar valued state variables as described above numerically, one needs to build a finitely computable numerical representation. This can be achieved very simply for models consisting only of integral and pointwise nonlinear composition maps by using a discretized version of the function valued state variables and a finite quadrature representation of the integrals, reducing evaluation of the functions defining the equilibrium conditions at a potentially infinite set of points to evaluation at a finite set and linearization to the calculation of a finite Jacobian matrix, which can be performed automatically by automatic differentiation software.[6] A challenge that one faces when doing this is evaluating and guaranteeing the accuracy of the discretized representations and solutions. As the exact and the discretized functions do not live in the same space, this requires putting the objects on a comparable footing.

A simple way to ensure comparability is to map the discrete points to functions through an interpolation procedure.[7] Beyond interpretability and amenability to quantitative guarantees, this approach has the substantial advantage that, especially when a model is defined in terms of smooth functions, one can attain quite high accuracy using only a small number of points. While in principle such an approach might require using a construction which requires repeated mapping back and forth between spaces in order to construct approximate solutions, for models built from pointwise nonlinear composition and integration only this can be reduced to a small number of post-processing steps, defined as simple matrix multiplies, which can be conducted automatically. The

[6] In the numerical experiments, this is performed in the environment Juliadiff [Revels et al., 2016], though any environment, such as Autograd [Maclaurin et al., 2015], that permits calculation of Jacobians by automatic differentiation through matrix operations could be used here.

[7] One might instead consider simply using a criterion which allows discrete and continuous objects to converge to each other, such as a weak convergence criterion; see Chatelin [2011] for an overview of such methods and the associated challenges.

model entered by the user can be defined using only discrete representations. The idea behind this equivalence is that for a linear integral operator, it is equivalent to interpolate the argument and output functions and to interpolate the kernel function of the integral operator. By composing guarantees for the accuracy of interpolation for each of the component steps, one can ensure the accuracy of the kernel representation, and then, using the accuracy guarantees for linearized solutions in terms of the approximation error in the derivatives of the equilibrium conditions from Childers [2018], ensure the accuracy of the solutions themselves.

Interpolation

In order to define a procedure based on interpolating the derivatives, I set out some preliminary objects including the class of interpolation methods that can be used. I will require the interpolation maps to satisfy a number of properties in order to provide accurate representations for the models of interest.

Definition 1. A (linear) interpolation scheme is a sequence of tuples $\{\mathcal{T}_K, M_K, \Phi_K\}$ consisting of a sequence of point sets $\mathcal{T}_K = \{s_k\}_{k=1}^K \in \mathcal{S}$ (not necessarily nested) and linear maps $M_K \in \mathbb{R}^{K\times K}$ such that for any input function $f(.) : \mathcal{S} \to \mathbb{R}$, letting $\vec{f} = \{f(s_k)\}_{k=1}^K$ and $\hat{f} = M_K\vec{f}$, for some set of basis functions $\Phi_K = \{\phi_k(s)\}_{k=1}^K$, produce an approximating function $\hat{f}_K(s) := \sum_{k=1}^K \hat{f}_k\phi_k(s)$.

Examples of interpolation schemes are polynomial interpolation at the Chebyshev points or otherwise, cubic spline interpolation, Gaussian process interpolation, and so on. See Appendix (A) for a list and discussion. In principle, function approximations which do not satisfy $\hat{f}(s_k) = f(s_k)$ for all k could be used with no loss in what follows, though for deterministic solution such methods are not needed.

The feature that will be needed for an interpolation scheme is that for appropriately smooth input functions, it produces output functions which are accurate in sup norm, and moreover, that for functions on $\mathcal{S}\times\mathcal{S}$ in an appropriate smoothness class, applying it argument by argument likewise produces sup norm accurate approximations. A *tensor product interpolation* takes an input function $f(.,.) : \mathcal{S}\times\mathcal{S} \to \mathbb{R}$ and given interpolation scheme $\{\mathcal{T}_K, M_K, \Phi_K\}$ produces the matrix $\vec{f} \in \mathbb{R}^{K\times K}$ with elements $[\vec{f}]_{ij} = f(s_i, s_j)$ and maps it into a matrix of basis coefficients $\hat{f} = M_K\vec{f}M_K^*$ producing an approximating function $\hat{f}_K(s,t) = \sum_{i=1}^K\sum_{j=1}^K \hat{f}_{ij}\phi_i(s)\phi_j(t)$

Definition 2. An interpolation scheme $\{\mathcal{T}_K, M_K, \Phi_K\}$ is *sup norm accurate* at rate ϵ_K for function class $\mathcal{F}$, a subset of functions $f(.) : \mathcal{S} \to \mathbb{R}$, if for any input function $f \in \mathcal{F}$, $\left\|\hat{f}_K(s) - f(s)\right\|_\infty \leq \epsilon_K$ for some sequence $\epsilon_K \to 0$ as $K \to \infty$. A tensor product interpolation based on $\{\mathcal{T}_K, M_K, \Phi_K\}$ is *sup norm accurate* at rate ζ_K for function class $\mathcal{F}^2$, a subset of functions $f(.,.)$ $\mathcal{S}\times\mathcal{S} \to \mathbb{R}$, if for any input function $f \in \mathcal{F}^2$, $\left\|\hat{f}_K(s,t) - f(s,t)\right\|_\infty \leq \zeta_K$ for some sequence $\zeta_K \to 0$ as $K \to \infty$.

For many standard interpolation schemes, including B-splines and Chebyshev polynomials, the default choice of basis functions may fail to be orthonormal. In these cases, one may construct an equivalent basis with identical approximation properties through orthonormalization. In such cases, M_K can be replaced by $J_K M_K$ and $\Phi_K = \{\phi_k(s)\}_{k=1}^K$ by $\Psi_K = \{\psi_k(s)\}_{k=1}^K$ for an equivalent interpolation matrix and orthonormal family. Furthermore, for variables which live in the space of densities, which must inegrate to 1, relevant perturbations reside in L_0^2 the subspace of square integrable functions which integrate to 0. Restriction of an interpolation scheme to operate only on this subspace can be achieved by orthogonalization of a standard basis with respect to the constant function, which can be achieved by replacing the interpolation matrix $M_K \in \mathbb{R}^{K-1\times K}$ with the modifed $S_K M_K \in \mathbb{R}^{K-1\times K}$ (or $S_K J_K M_K$ if both standardization and orthonormalization are needed). Generic procedures for constructing both classes of transform are reviewed in Appendix (A.1). In further discussions, interpolation schemes $(\mathcal{T}_K, M_K, \Phi_K)$ will always be assumed to refer to schemes which have been orthonormalized and, for density-valued variables, standardized in this way.

Given an interpolation scheme and, if needed, a map which converts the output of the interpolation scheme into coefficients with respect to an orthonormal basis and possibly normalizes as well if the functions are constrained, an integral kernel operator $\int k(t,s)[.]ds$ with $k(t,s): \mathcal{S}\times\mathcal{S}\to\mathbb{R}$ may be approximated by a representation of the operator in terms of this basis by mapping the coefficients of the tensor product interpolation of the kernel to the coefficients with respect to an orthonormal basis. This produces a representation $\sum_{i=1}^K \sum_{j=1}^K \theta_{i,j} \langle \psi_j, . \rangle \psi_i$, where, if $[K]_{i,j} = K(t_i, s_j)$, $\theta_{i,j} = [J_K M_K K M_K^* J_K^*]_{ij}$ if not normalized and $\theta_{i,j} = [S_K J_K M_K K M_K^* J_K^* S_K^*]_{ij}$ if normalized. Furthermore, supposing that the tensor product interpolation scheme is sup norm accurate, an application of Young's inequality (and an operator norm bound on S_K if applied) implies that if $k(t,s) \in \mathcal{F}^2$, $\left\| \int k(t,s)[.]ds - \sum_{i=1}^K \sum_{j=1}^K \theta_{i,j} \langle \psi_j, . \rangle \psi_i \right\|_{op}$ converges to 0 as $K \to \infty$. Sufficient conditions will be given below such that this is the case. While this transformation may be used directly to construct a basis representation of the functional derivatives satisfying the conditions of Theorem 1 in Childers [2018] and so suitable for application of a linear rational expectations solver, this requires constructing functional derivatives before taking the appropriate transforms.

A key restriction is that the above mapping result applies only to operators which take kernel integral operator form. In general, even when a model is defined only in terms of kernel integral and piecewise nonlinearity maps, its functional derivatives need not take only this form. Certain functional derivatives will instead take the form of identity operators, or multiplication of an input function by another function. This latter type can be detected and solved out automatically;[8] the former require special treatment. The numerical derivatives

[8] Cases where this cannot be solved out may also exist; these violate Condition (4) below and create problems for the algorithm and indeed any method based on the guarantees of Childers [2018]. Current implementations return a warning in this case.

in the case of an identity map will produce a finite identity matrix. In general, applying an interpolation scheme to an identity matrix will not produce an accurate representation of an identity map on a space of coefficients.[9] This is easily resolved, however, as an identity matrix, interpreted as a map over a set of coefficients of an orthonormal basis, is a valid representation of an identity operator, and so the solution is to only apply the interpolation transforms to the model components which take the form of kernel integral operators. These can be determined, even without taking derivatives, by the structure of the model. By detecting the appropriate components and transform them, I can now define a general set of conditions and an algorithm for model solution based on interpolation schemes described here.

A very general class of operators

In order to provide a procedure applicable to the widest feasible class of economic models amenable to a linearization procedure based on pointwise differentiation, it will be helpful to introduce notation compatible with a variety of models. In essence, the models which can be handled by a procedure in this class consist of those derived from repeated composition of linear integral operators and pointwise nonlinearities. In the case where functions of interest live on a finite support, this is a class which includes (most) deep neural network architectures; including the generalization of neural networks to infinite-dimensional input and output data as examined by Guss [2017].

The following conditions define the class of model equilibrium conditions in terms of repeated composition of integral and pointwise composition operators along a structured graph. Condition (3) defines the components of this graph in a way which ensures the input and output functions have the right arguments. Condition (4) imposes primitive conditions on the graph topology which ensure that the derivatives take the form of Fredholm integral equations of the second kind by making sure that all paths from input to output either consist of an identity map or contain a kernel integral operator. Notation is introduced to mark distinctions between these classes of path, which will need to be treated differently when interpolating the discretized model. Given a model following these conditions, an algorithm (Algorithm 1) is introduced which constructs approximate functional derivatives through discretization, differentiation, and interpolation along the graph. Smoothness conditions on the steady state functions which are composed to produce the equilibrium conditions (Condition (5)) and their derivatives (Condition (6)) are provided which ensure that the steady state functions and the resulting functional derivatives are smooth enough to be approximated with (asymptotically) negligible error, with rates quantified in Lemmas (7) and (9), respectively. As a result, under the approximation sta-

[9] An exception to this is when the maps $J_K M_K$ are unitary, so $J_K M_K (J_K M_K)^* = I_k$. This does occur, possibly up to a rescaling, for certain interpolation schemes, in particular some classes of wavelet and histogram methods: see Appendix (A). In these cases, a somewhat simpler version of the procedure developed here can be used. This is described in a companion paper, Childers and Dogra [2018].

bility conditions for linear rational expectations solution algorithms provided in Childers [2018] (which include generalizations of the Blanchard Kahn local existence and uniqueness conditions for the model solution which are *not* guaranteed by these primitive conditions), the resulting linearized solution will also have negligible error.

The class of permissible operators $\mathcal{F}_\ell$ describing each equation $\ell = 1 \dots d_2$ will be constructed iteratively as the composition of simpler operators organized along a directed acyclic computational graph, with inputs given by the function-valued state variables $\{g_j(.)\}_{j=1}^{2d_2}$. While most practical models of interest are likely to correspond to a fairly shallow computational graph, the use of such a formalism facilitates construction of functional derivatives through the chain rule in a principled fashion, and allows, if needed, a highly expressive class for approximation of nonlinear operators, building on universal approximation results for neural networks.

Condition 3. Suppose that $\mathcal{S} = \times_{i=1}^{q} \mathcal{S}_i$, where i indexes a mutually exclusive and exhaustive partition of $q \leq d_1$ subsets of dimensions of $\mathcal{S}$ and each element $j = 1 \dots 2d_2$ of the set of state variables $[g(s)]_j = g_j(s_{[j]})$ depends on a subset $[j] \subseteq \{1 \dots q\}$ of these subsets of dimensions, denoted $\mathcal{S}_{[j]} = \underset{i \in [j]}{\times} \mathcal{S}_i$. In total, computations may reference up to $\chi + 1$ possible arguments, indexed by $i \in \{1 \dots \chi + 1\}$, where the first q correspond to $\mathcal{S}_i$, $i \in \{1 \dots q\}$, additional arguments $i \in \{q + 1 \dots \chi\}$, $\chi \geq q$, live on some spaces $\mathcal{S}_i$, where $\mathcal{S}_i$ for $i > q$ is some compact subset of a finite-dimensional Euclidean space, and argument $\chi + 1$ is t living on $\mathcal{S}_{[\ell^o]}$ for some subset $[\ell^o] \subset \{1 \dots q\}$.

For each condition $\ell = 1 \dots d_2$, the equilibrium condition operator $\mathcal{F}^\ell$ is described by a directed, acyclic, connected computational graph with P^ℓ nodes. Nodes $p^\ell \in \{1 \dots P^\ell\}$ in the graph each produce as output a function $n^{p^\ell} : \mathcal{S}_{[p^\ell]} \to \mathbb{R}$, where $[p^\ell] \subseteq \{1 \dots \chi + 1\}$ is some subset of possible arguments, and may take as input the functions produced by parent nodes $par(p^\ell)$. There is one output node n^{P_ℓ} (with only incoming edges) whose output is a function only of $t \in \mathcal{S}_{[\ell^o]}$. For each input function $j = 1 \dots 2d_2$, there is one input node with index $p^\ell = j$ with no parents and output $n^j = g_j(s_{[j]})$. In addition, there may be $o^\ell \leq 2d_2$ additional input nodes, each corresponding to a separate input function $j(p^\ell) \in \{1 \dots 2d_2\}$ such that $\mathcal{S}_{[j(p^\ell)]} \cong \mathcal{S}_{[\ell^o]}$ and which produce output $n^{p_\ell} = g_{j(p^\ell)}(t) : \mathcal{S}_{[\ell^o]} \to \mathbb{R}$.

All intermediate nodes consist of two components, a nonlinear map and a subsequently applied linear map. The composition of these two maps applied to the output of all ingoing edges generates the function n^{p^ℓ}. The nonlinear map corresponding to index p^ℓ with parents $par(p^\ell) \subset \{1 \dots P^\ell\}$ is associated with a function $f^{p^\ell}(s_{[p_I^\ell]}, (.)) : \mathcal{S}_{[p_I^\ell]} \times \mathbb{R}^{|par(p^\ell)|} \to \mathbb{R}$ for some subset $[p_I^\ell] \subset \{1 \dots \chi + 1\}$ which is composed with the output of its parent nodes to produce output $f^{p^\ell}(s_{[p_I^\ell]}, n^{[par(p^\ell)]}(s_{[par(p^\ell)]})) : \mathcal{S}_{[p_n^\ell]} \to \mathbb{R}$ for $[p_n^\ell] = [p_I^\ell] \cup \underset{p \in par(p^\ell)}{\cup} [p]$. Within each node, this function is then composed with a linear map, corresponding to a linear operator taking one of two forms:

either an identity map I or an integral operator $\int k^{p^\ell}(s_{[p_a^\ell]})[.]ds_{[p_b^\ell]}$ for some subsets $[p_a^\ell] \subset \{1 \dots \chi\}$, $[p_b^\ell] \subset ([p_a^\ell] \cup [p_n^\ell])/\{\chi+1\}$. In the identity case, the resulting output takes form $n^{p^\ell}(s_{[p^\ell]}) = f^{p^\ell}(s_{[p_I^\ell]}, n^{[par(p^\ell)]}(s_{[par(p^\ell)]}))$ for $[p^\ell] = [p_n^\ell]$. In the integral operator case, the node output takes form $n^{p^\ell}(s_{[p^\ell]}) = \int k^{p^\ell}(s_{[p_a^\ell]}) f^{p^\ell}(s_{[p_I^\ell]}, n^{[par(p^\ell)]}(s_{[par(p^\ell)]}))ds_{[p_b^\ell]}$ where $[p^\ell] = ([p_a^\ell] \cup [p_n^\ell])/[p_b^\ell]$. The final output node P^ℓ takes the same structure as all intermediate nodes, with the restriction that $[P^\ell] = \{\chi+1\}$, and the final output of equilibrium condition ℓ is given by $\mathcal{F}^\ell(\{g_j(.)\}_{j=1}^{2d_2})(t) = n^{P_\ell}(t)$.

The above condition generates a broad class of operators whose derivatives may be approximately computed in an automated fashion using only pointwise information. In order to apply solution procedures, I impose additional regularity conditions on the operators. Condition (a) ensures maps and derivatives are well-defined, while condition (b) ensures that the operator, after suitable transformation, produces derivatives which satisfy the compact plus identity representation condition in Childers [2018].

Condition 4. (a) All nonlinear maps f^{p^ℓ} are bounded, measurable, and differentiable in all arguments. Further, both f^{p^ℓ} and the derivative of $f^{p^\ell}()$ with respect to any parent node should be Hölder continuous with Hölder exponent $\alpha^{p^\ell} > 0$ in each argument corresponding to a parent node, uniformly in all other arguments. Similarly, all kernel functions $k^{p^\ell}(s_{[p_a^\ell]})$ should be bounded and measurable.

(b) For all equilibrium conditions $\ell = 1 \dots d_2$, the operator $\mathcal{F}^\ell$ satisfies one of the following two conditions.

(i) $o^\ell \leq 1$, so that there is at most one input node $n^{p^\ell(out)}$, $p^\ell(out) = 2d_2 + 1$ such that $n^{p^\ell(out)} = g_{j(p^\ell(out))}(t)$ where $\mathcal{S}_{[j(p^\ell(out))]} \cong \mathcal{S}_{[\ell^o]}$, or

(ii) For all input nodes n^{p^ℓ} such that $p^\ell \in (2d_2+1) \dots (2d_2+o^\ell)$, so that $n^{p_\ell} = g_{j(p^\ell)}(t)$ where $\mathcal{S}_{[j(p^\ell)]} \cong \mathcal{S}_{[\ell^o]}$, all descendant nodes have a nonlinear component $f(.)$ which is linear in any argument such that n^{p_ℓ} is an ancestor and linear component given by the identity map.

Due to the generality of directed acyclic graph structure, this class of operators includes as a subset Hammerstein integral operators and sums thereof, the Deep Functional Networks of Guss [2017], and many more general functions, which do not necessarily take a tree structure or have recursive representation. By restricting to acyclic graphs, it does rule out recursive maps, though any finite set of iterations can be generated by passing through a long enough path of identical nodes. The Hölder condition is a sufficient condition to ensure that not only will a finite approximation of each node converge, but that passing the approximation error through a computational graph of arbitrary finite depth will not cause the approximation error to explode. It is satisfied if each function is Hölder of order $\alpha^{p^\ell}+1$ over all of its arguments, though this is not a necessary condition.

The true functional derivatives to be approximated can be computed by accumulating along the directed computational graph either through forward or reverse mode (or through a combination thereof), as functional derivatives follow a standard chain rule. While the task of computing the derivative of each output $\mathcal{F}^\ell$ with respect to $2d_2 + o^\ell$ input functions suggests advantages for reverse mode, this depends on the topology of the computational graph and in computational experiments forward mode appears to outperform.

To construct the partial derivatives at each node, first note that the derivative of the linear component of any node is equal to the component itself. For nonlinear components, the derivative with respect to any ingoing node $i \in par(p^\ell)$ at a value $n^{[par(p^\ell)]*}$ of its inputs is given by the multiplication operator $\frac{d}{dn^i} f^{p_\ell}(s_{[p_I^\ell]}, n^{[par(p^\ell)]*}(s_{[par(p^\ell)]})) \cdot [.]$ where the derivative is the ordinary scalar derivative of the function f^{p^ℓ} with respect to the argument corresponding to input n^i.

Approximation of functional derivatives of the above form can be performed by representing the functions generating all operators using tensor product interpolation schemes. While a variety of methods are available for approximating the integrals in these operators, I will mainly focus on a particular choice with both desirable ease of use and accuracy properties. In particular, all integrals in the discretized equilibrium conditions may be represented by the exact quadrature scheme corresponding to the interpolation scheme. The resulting nonlinear map then takes as input $2d_2 + o^\ell$ vectors corresponding to the value of the input functions at the appropriate grid points, and produces as output a vector corresponding to the (approximate) value of the output function $\mathcal{F}^\ell(\{\overrightarrow{g}_j\}_{j=1}^{2d}, \{\overrightarrow{g}_{j(p^\ell)}\}_{p^\ell=2d_2+1}^{2d_2+o^\ell})(t)$ over a grid of points on $\mathcal{S}_{[\ell^o]}$. This map between finite dimensional vector spaces can then be differentiated according to the same computational graph as the exact operator by standard automatic differentiation software over vectors to produce a set of Jacobian matrices. These matrices can then be transformed into representations of functional derivatives by applying a few simple matrix transforms and then the appropriate interpolation maps to produce matrices of coefficients, which can then be put through standard rational expectations solution software.

To construct this procedure, let $\{\mathcal{T}_z, M_z, \Phi_z\}_{z=1}^{\chi}$ be a set of interpolation schemes with respective cardinalities K_z for functions on domains $\{\mathcal{S}_z\}_{z=1}^{\chi}$ respectively. Note that a tensor product interpolation scheme over functions on a subset $[p] \subseteq \{1 \dots \chi\}$ with cardinality $K_{[p]} = \Pi_{z\in[p]} K_z$ can be constructed as $\mathcal{Q}_{[p]} := \{\times_{z\in[p]}\mathcal{T}_z, \otimes_{z\in[p]} M_z, \otimes_{z\in[p]}\Phi_z\}$, where $M_{[p]} := \otimes_{z\in[p]} M_z$ indicates the Kronecker product of the matrices M_z and $\otimes_{z\in[p]}\Phi_z$ consists of the set of tensor products over the sets Φ_z of basis functions. Without loss of generality, assume the basis sets Φ_z are orthonormal: if this is not true, M_z, Φ_z may be replaced by orthonormalized versions as constructed by Algorithm (2). Denote by $\mathcal{Q}_{[\chi+1]} := \{\mathcal{T}_{\chi+1}, M_{\chi+1}, \Phi_{\chi+1}\} := \{\times_{z\in[\ell^o]}\mathcal{T}_z, \otimes_{z\in[\ell^o]} M_z, \otimes_{z\in[\ell^o]}\Phi_z\}$, where I have abused notation as this object depends on ℓ; however the use will be clear from context. I will assume that each individual scheme is sup norm accurate, as are all tensor product schemes used, for functions in appropriate function classes.

Primitive conditions for this will depend on the scheme and the function classes used. For example, tensor product spline or wavelet approaches over Hölder classes or, if some $\mathcal{S}_z$ are multidimensional, tensor products of sparse grid spline interpolation over spaces of continuous functions with bounded mixed partial derivatives will ensure consistency over any subset of dimensions.

An interpolation scheme $\{\mathcal{T}_z, M_z, \Phi_z\}$ induces an exact quadrature scheme with abcissas (evaluation points) $\mathcal{T}_z$ and weights $\pi^z = (q_z' M_z)'$, where the j^{th} element of q_z equals $\int_{\mathcal{S}_z} \phi_j(s_z) ds_z$, the integral of the j^{th} basis function. It is easily seen that the quadrature scheme corresponding to a sup norm accurate interpolation scheme over a bounded domain converges over the same function class at a rate identical to the sup norm approximation rate. For tensor product quadrature over functions on a subset $[p] \subseteq \{1 \dots \chi\}$, denote $\pi^{[p]} = \otimes_{z \in [p]} \pi^z$ the length $K_{[p]}$ vector of weights corresponding to abcissas $\times_{z \in [p]} \mathcal{T}_z$, and a corresponding result holds. Beyond the accuracy properties, schemes of this form, unlike Gaussian quadrature, have the advantage that they use the same function evaluations as used to construct an interpolation. In practice, it is the latter property which is necessary for the algorithm; non-exact schemes which use the same grid points as abcissas can be used, and will result in similar guarantees so long as they are sufficiently accurate. By representing all equilibrium conditions using the interpolation scheme corresponding to the variables contained in each state and using these quadrature schemes to ensure additional dimensions get integrated out, one can construct an algorithm which accounts for the different arguments of each state, introduced as Algorithm 1.

Definition. Let $k^{p^\ell}(s_{[p_a^\ell]}) \in \mathbb{R}^{K_{[p_a^\ell]} \times 1}$ be the kernel of the integral operator of node p^ℓ evaluated at $K_{[p_a^\ell]}$ points, and let $\hat{\Pi}^{[p_a^\ell] \cup [p_n^\ell]} \in \mathbb{R}^{K_{[p_a^\ell] \cup [p_n^\ell]} / K_{[p_b^\ell]} \times K_{[p_a^\ell] \cup [p_n^\ell]}}$ be $\hat{\Pi}^{[p_a^\ell] \cup [p_n^\ell]} := I_{K_{[p_a^\ell] \cup [p_n^\ell]} / K_{[p_b^\ell]}} \otimes \pi^{[p_b^\ell]\prime}$ for $\pi^{[p_b^\ell]}$ a vector of quadrature weights over the abcissas $\times_{z \in [p_b^\ell]} \mathcal{T}_z$. Then

$$L^{p^\ell} := \hat{\Pi}^{[p_a^\ell] \cup [p_n^\ell]} \mathrm{Diag}(k^{p^\ell}(s_{[p_a^\ell]}) \otimes 1_{K_{[p_a^\ell] \cup [p_n^\ell] / [p_a^\ell]}}) (1_{K_{[p_a^\ell] \cup [p_n^\ell] / [p_n^\ell]}} \otimes I_{K_{[p_n^\ell]}}).$$

This representation of the integral operator $\int k^{p^\ell}(s_{[p_a^\ell]})[.] ds_{[p_b^\ell]}$ serves the sole purpose of reshaping the discretized operator with integrals approximately evaluated by quadrature into matrix form, so that its derivative may be evaluated efficiently (viz, by linearity, as the integral itself). Alternately, in a conceptually simpler but more memory-intensive representation, this operator and all intermediate functions could be expressed in terms of tensors of order (up to) $\chi + 1$, with all operations except integration then expressible as pointwise operations and integration through summing over respective arguments. This scheme would be numerically equivalent, and likely easier to code, but is extremely memory inefficient, with each tensor having dimension and so memory cost exponential in χ.

The application of Algorithm (1) yields a consistent approximation of the functional derivatives of operator $\mathcal{F}$ under mild conditions on the model and the

Algorithm 1 Construction of spectral approximation over a graph

Input: Equilibrium conditions $\mathcal{F}^{\ell}(g(.))(t)$ as in Condition (3), approximate steady state functions $\tilde{g}^{*}(t)$, and interpolation schemes $\{\mathcal{T}_z, M_z, \Phi_z\}_{z=1}^{\chi}$
Output: Approximate solutions $\tilde{h}_x^K$, $\tilde{g}_x^K$

1. If $\{\Phi_z\}_{z=1}^{\chi}$ not each orthonormal, apply Algorithm (2) to orthonormalize all interpolation schemes

2. Build vectors of input variables

   (a) Construct input vectors $\{\overrightarrow{g}_j\}_{j=1}^{2d_2}$ as $x := \{\{g_j(s_{[j]i})\}_{i=1}^{K_{[j]}}\}_{j=1}^{d_x}$, $y := \{\{g_j(s_{[j]i})\}_{i=1}^{K_{[j]}}\}_{j=d_x+1}^{d_2}$, $x' := \{\{g_j(s_{[j]i})\}_{i=1}^{K_{[j]}}\}_{j=d_2+1}^{d_2+d_x}$, and $y' := \{\{g_j(s_{[j]i})\}_{i=1}^{K_{[j]}}\}_{j=d_2+d_x+1}^{2d_2}$, where $\{s_{[j]i}\}_{i=1}^{K_{[j]}} = \times_{z\in[j]}\mathcal{T}_z$

   (b) Construct $\{\overrightarrow{g}_{j(p^\ell)}\}_{p^\ell=2d_2+1}^{2d_2+o^\ell}$ as $\{\{g_{j(p^\ell)}(t_{[j(p^\ell)]i})\}_{i=1}^{K_{[j(p^\ell)]}}\}_{p^\ell=2d_2+1}^{2d_2+o^\ell}$ where $\{t_{[j(p^\ell)]i}\}_{i=1}^{K_{[j(p^\ell)]}} = \times_{z\in[j(p^\ell)]}\mathcal{T}_z$

   (c) For $p^\ell = 1\dots 2d_2+o^\ell$, define nodes $\overrightarrow{n}^{p^\ell}(\overrightarrow{g}_{j(p^\ell)}) := \overrightarrow{g}_{j(p^\ell)}$

3. For each node $p^\ell = 2d_2+o^\ell+1\dots P^\ell$, construct node representation $\overrightarrow{n}^{p^\ell}(\{\overrightarrow{n}^{p}\}_{p\in par(p^\ell)}) \underset{p\in par(p^\ell)}{\times} \mathbb{R}^{K_{[p]}} \to \mathbb{R}^{K_{[p^\ell]}}$ by composing

   (a) Nonlinear component $\overrightarrow{f}^{p^\ell}()$ as map $\underset{p\in par(p^\ell)}{\times} \mathbb{R}^{K_{[p]}} \to \mathbb{R}^{K_{[p_n^\ell]}}$ with inputs $\{\overrightarrow{n}^{p}\}_{p\in par(p^\ell)}$ and values $\{f^{p^\ell}(s_{[p_I^\ell]i}, n^{[par(p^\ell)]}(s_{[par(p^\ell)]i}))\}_{i=1}^{K_{[p_n^\ell]}}$

   (b) If node p^ℓ has linear map I, apply matrix $I_{K_{[p_n^\ell]}}$. If linear map is of form $\int k^{p^\ell}(s_{[p_a^\ell]})[.]ds_{[p_b^\ell]}$, apply $K_{[p^\ell]} \times K_{[p_n^\ell]}$ matrix L^{p^ℓ} specified below.

4. $\forall \ell = 1\dots d_2$, apply automatic differentiation along the computational graph with respect to $\{\overrightarrow{g}_j\}_{j=1}^{2d_2}$ to build matrices $\{\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}}\}_{j=1}^{2d_2}$ and to $\{\overrightarrow{g}_{j(p^\ell)}\}_{p^\ell=2d_2+1}^{2d_2+o^\ell}$ to build matrices $\{\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^\ell)(out)}}\}_{p^\ell=2d_2+1}^{2d_2+o^\ell}$ around $\{\overrightarrow{g}^{*}_{j(p^\ell)}\}_{j=1}^{2d_2+o^\ell}$, the (approximate) steady state values of $\{\overrightarrow{g}_{j(p^\ell)}\}_{j=1}^{2d_2+o^\ell}$.

5. $\forall \ell = 1\dots d_2$, if ℓ satisfies Condition (4)(b)(ii), $\forall j = 1\dots 2d_2$ represent the (ℓ, j) block of the equilibrium equations as $\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(out)}} + M_{[\ell^o]}\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}}(\Pi^{[j]})^{-1}M^{*}_{[j]}$, while if ℓ satisfies Condition (4)(b)(i), $\forall j = 1\dots 2d_2$ represent the (ℓ, j) block of the equilibrium equations as $(\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^\ell(out))}})^{-1}\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(out)}} + M_{[\ell^o]}(\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^\ell(out))}})^{-1}\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}}(\Pi^{[j]})^{-1}M^{*}_{[j]}$

6. Apply standard rational expectations solution algorithm as in Schmitt-Grohé and Uribe [2004] to the system of equations to obtain solution matrices g_x^K, h_x^K of coefficient representations of solutions which result in maps $\tilde{g}_x^K = \sum_i \sum_j [g_x^K]_{ij} \langle \varphi_j(.), . \rangle \phi_i(.)$ and $\tilde{h}_x^K = \sum_{i=1} \sum_{j=1} [h_x^K]_{ij} \langle \varphi_j(.), . \rangle \varphi_i(.)$ for $\varphi \in \otimes_{z\in\{1\dots d_x\}}\Phi_z$, $\phi \in \otimes_{z\in\{d_x+1\dots d_2\}}\Phi_z$

interpolation schemes used. Due to the nature of the functions to be approximated, it is necessary to apply approximation procedures and conditions which maintain accuracy when interpolation and function composition are repeatedly composed. While weaker conditions are likely possible, the following set of restrictions provides a simple set of sufficient conditions for these approximations to be valid.

Condition 5. (i) For all ℓ, for all $p^\ell = 1 \dots P^\ell$, for each $p \in par(p^\ell)$ let $f^{p^\ell}(s_{[p_I^\ell]}, (.))$ be Hölder continuous with exponent $\alpha^{p^\ell(p)}$ in input $n^p(.)$ (considered as a scalar) uniformly over all inputs other than p

(ii) For all ℓ, for all $p^\ell = 1 \dots P^\ell$, let $\mathcal{G}^{[p_a^\ell] \cup [p_n^\ell]} : \mathcal{S}_{[p_a^\ell] \cup [p_n^\ell]} \to \mathbb{R}$ and $\mathcal{G}^{[p^\ell]} : \mathcal{S}_{[p^\ell]} \to \mathbb{R}$ be classes of bounded functions on $\mathcal{S}_{[p_a^\ell] \cup [p_n^\ell]}$ and $\mathcal{S}_{[p^\ell]}$, respectively, such that for all ℓ, for all $p^\ell = 1 \dots P^\ell$ $k^{p^\ell}(s_{[p_a^\ell]}) f^{p^\ell}(s_{[p_I^\ell]}, (.)) : \underset{p \in par(p^\ell)}{\times} \mathcal{G}^{[p]} \to \mathcal{G}^{[p_a^\ell] \cup [p_n^\ell]}$ and, if the linear component of node p^ℓ is an integral operator $\int [] ds_{[p_b^\ell]} : \mathcal{G}^{[p_a^\ell] \cup [p_n^\ell]} \to \mathcal{G}^{[p^\ell]}$, otherwise $\mathcal{G}^{[p^\ell]} := \mathcal{G}^{[p_n^\ell]}$, i.e., such that the application of each node maps to an appropriate function class.

(iii) Let steady state state variables $g_j^*(.) \in \mathcal{G}^{[j]}$

(iv) Given tensor product interpolation schemes $\mathcal{Q}_{[p_a^\ell] \cup [p_n^\ell]}$, for all ℓ, for all $p^\ell = 1 \dots P^\ell$, let $\hat{\mathcal{G}}^{[p_a^\ell] \cup [p_n^\ell]} : \mathcal{S}_{[p_a^\ell] \cup [p_n^\ell]} \to \mathbb{R}$ and $\hat{\mathcal{G}}^{[p^\ell]} : \mathcal{S}_{[p^\ell]} \to \mathbb{R}$ be classes of bounded functions on $\mathcal{S}_{[p_a^\ell] \cup [p_n^\ell]}$ and $\mathcal{S}_{[p^\ell]}$, respectively, such that for all ℓ, for $p^\ell = 1 \dots 2d_2 + o_\ell$, $\tilde{g}^*_{j(p^\ell)} \in \hat{\mathcal{G}}^{[p^\ell]}$, and for all $p^\ell = 1 \dots P^\ell$, (a) $k^{p^\ell}(s_{[p_a^\ell]}) f^{p^\ell}(s_{[p_I^\ell]}, (.)) : \underset{p \in par(p^\ell)}{\times} \hat{\mathcal{G}}^{[p]} \to \hat{\mathcal{G}}^{[p_a^\ell] \cup [p_n^\ell]}$, and (b) if the linear component of node p^ℓ is an integral operator, the quadrature scheme $\pi^{[p_b^\ell]}$ corresponding to integral $\int [] ds_{[p_b^\ell]}$ maps $\hat{\mathcal{G}}^{[p_a^\ell] \cup [p_n^\ell]} \to \hat{\mathcal{G}}^{[p^\ell]}$, otherwise $\hat{\mathcal{G}}^{[p^\ell]} := \hat{\mathcal{G}}^{[p_n^\ell]}$, i.e., that each approximated node likewise maps to an appropriate function class

(v) Let $\hat{\mathcal{G}}^{[p_b^\ell]} : \mathcal{S}_{[p_b^\ell]} \to \mathbb{R}$ be a class of functions such that $\forall f(s_{([p_a^\ell] \cup [p_n^\ell])/[p_b^\ell]}, s_{[p_b^\ell]}) \in \hat{\mathcal{G}}^{[p_a^\ell] \cup [p_n^\ell]}$, $\forall s_{([p_a^\ell] \cup [p_n^\ell])/[p_b^\ell]} \in \mathcal{S}_{([p_a^\ell] \cup [p_n^\ell])/[p_b^\ell]}$, $f(s_{([p_a^\ell] \cup [p_n^\ell])/[p_b^\ell]}, .) \in \hat{\mathcal{G}}^{[p_b^\ell]}$

(vi) For all ℓ, for $j = 1 \dots 2d_2 + o^\ell$, let the approximations of each steady state function converge at rate $\zeta^{p^\ell}(K)$, i.e. $\left\| \tilde{g}^*_{j(p^\ell)} - g^*_{j(p^\ell)} \right\|_\infty \leq \zeta^{p^\ell}(K)$, and for all ℓ, for all p^ℓ, (i) let tensor product interpolation schemes $\mathcal{Q}_{[p^\ell]}$ be sup norm accurate over classes $\hat{\mathcal{G}}^{[p^\ell]}$ at rates $\zeta^{[p^\ell]}(K_{[p^\ell]})$, (ii) let tensor product interpolation schemes $\mathcal{Q}_{[p_b^\ell]}$ be sup norm accurate over classes $\hat{\mathcal{G}}^{[p_b^\ell]}$ at rates $\zeta^{[p_b^\ell]}(K_{[p_b^\ell]})$

Remark. Regarding condition (5), the function classes invoked in this theorem and the resulting rates will depend on the smoothness assumed for the structural objects, how this is preserved under composition, multiplication, integration, and quadrature, and the appropriateness of the interpolation schemes used over these classes. Note that Hölder continuity of f^{p^ℓ} and k^{p^ℓ} over all inputs guarantees condition (i) and also ensures that condition (ii) holds if function classes $\mathcal{G}^{[p^\ell]}$ are given by the Hölder classes $\Lambda^{\alpha_{[p^\ell]}}[\mathcal{S}_{[p^\ell]}]$ for appropriate exponents $\alpha_{[p^\ell]}$

guided by composition rules for Hölder functions. A useful sufficient condition is that if f^{p^ℓ} and k^{p^ℓ} have Hölder exponent greater than or equal to 1 for all ℓ, p, then $\alpha_{[p^\ell]} \geq 1$. Sufficient conditions for (iii) will depend on the properties of the full system: depending on the class, proof of this may proceed by the use of a fixed point theorem. Parts (iv) and (v) of this condition hold for a number of commonly used combinations of function classes and interpolation schemes, with rates depending on the degree of accuracy of the interpolation scheme over the induced classes. While (iv)(a) may generally be derived from the smoothness-preserving properties of integration, multiplication, and composition with smooth functions, holding, for example, over Hölder classes as in (ii), (iv)(b) requires that when replacing exact integration by numerical quadrature, as performed in the algorithm, the resulting function preserves a degree of smoothness. In the case of functions which are Hölder smooth in $s_{([p_a^\ell]\cup[p_n^\ell])/[p_b^\ell]}$ uniformly over $s_{[p_b^\ell]}$, as is true for, functions in, say, $\hat{\mathcal{G}}^{[p_a^\ell]\cup[p_n^\ell]} = \Lambda^\alpha(\mathcal{S}_{([p_a^\ell]\cup[p_n^\ell])})$ for any $\alpha > 0$, this is true because for any quadrature scheme with summable weights the output function remains in a Hölder ball in $s_{([p_a^\ell]\cup[p_n^\ell])/[p_b^\ell]}$ regardless of the order of integration, as follows from the characterization of Hölder classes as hyperrectangles in the space of Wavelet coefficients (see Johnstone [2015] Ch 4.7). This property likewise extends to other function classes requiring uniform bounds on derivatives, such as tensor products of $F_{d_z}^{\alpha_z}(\mathcal{S}_z)$ (see Appendix (A) for definition), as these are preserved under summable linear combinations. For other function classes which may not in general be preserved under infinite sums (including Sobolev and non-uniform Besov classes), this property is harder to guarantee, but may still hold under slightly stronger conditions on the integrand. Condition (v) is a "marginalization" condition which requires that fixing one argument of a regular function, it remains a regular function of the other arguments. This holds for a variety of regularity classes with uniform definitions: in particular, if $\hat{\mathcal{G}}^{[p_a^\ell]\cup[p_n^\ell]} = \Lambda^\alpha(\mathcal{S}_{([p_a^\ell]\cup[p_n^\ell])})$ for any $\alpha > 0$, $\hat{\mathcal{G}}^{[p_b^\ell]} = \Lambda^\alpha(\mathcal{S}_{[p_b^\ell]})$ may be used here. Tensor products of $F_{d_z}^{\alpha_z}(\mathcal{S}_z)$, along with any other condition imposing uniform bounds on partial derivatives likewise satisfy this property, with appropriate adjustments for the dimension of the function. Regarding the rates in part (vi), for intermediate nodes these may be derived from standard function approximation rates.

For the input functions at the steady state, bounds on $\left\|\sum_{k=1}^{K_{[j]}}(M_{K_{[j]}}\overrightarrow{\tilde{g}}^*_{j(p^\ell)})_k\phi_k - g^*_{j(p^\ell)}\right\|_\infty$ can be derived from the triangle inequality and Condition (5)(v) and (vi), which requires that the approximation of the steady state is reasonably accurate and that the interpolation scheme is sup norm accurate. As a wide variety of algorithms are used for steady state approximation, I leave only the high level condition that the approximation lies in a useful smoothness class and is uniformly accurate. Uniform accuracy can be guaranteed by a variety of fixed point algorithms, while smoothness of the approximation is a property of a variety of projection schemes using sufficiently regular basis functions, though it is worth noting that even if the true functions g_j^* lie in a class $\mathcal{G}^{[j]}$, uniformly accurate approximations may not remain in this class. For example, conver-

gence of values need not imply convergence of derivatives, or even existence of derivatives. Nevertheless, possibly by allowing $\hat{\mathcal{G}}^{[j]}$ to be less regular than $\mathcal{G}^{[j]}$, a reasonable degree of smoothness (and therefore, reasonably fast convergence rates) may still be attainable.

Condition 6. (i) For all ℓ, for all $p^\ell = 1 \dots P^\ell$, for each $p \in par(p^\ell)$ let the partial derivative $\frac{\partial}{dn^p} f^{p^\ell}(s_{[p_I^\ell]}, (.))$ of $f^{p^\ell}(s_{[p_I^\ell]}, (.))$ with respect to input $n^p(.)$ be Hölder continuous with exponent $\alpha_d^{p^\ell(p)}$ in input $n^p(.)$ (considered as a scalar) uniformly over all inputs other than p.

(ii) For all ℓ, for all $p^\ell = 1 \dots P^\ell$, let $\mathcal{G}_d^{[p_n^\ell]} : \mathcal{S}_{[p_n^\ell]} \to \mathbb{R}$, $\mathcal{G}_d^{[p_a^\ell] \cup [p_n^\ell]} : \mathcal{S}_{[p_a^\ell] \cup [p_n^\ell]} \to \mathbb{R}$ and $\mathcal{G}_d^{[p^\ell]} : \mathcal{S}_{[p^\ell]} \to \mathbb{R}$ be classes of bounded functions on $\mathcal{S}_{[p_n^\ell]}$, $\mathcal{S}_{[p_a^\ell] \cup [p_n^\ell]}$, and $\mathcal{S}_{[p^\ell]}$, respectively, such that for all ℓ, for all $p^\ell = 1 \dots P^\ell$ for each $q \in par(p^\ell)$, if the linear component of node p^ℓ is an integral operator, $k^{p^\ell}(s_{[p_a^\ell]}) \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, (.)) \cdot (.) : \underset{p \in par(p^\ell)}{\times} \mathcal{G}^{[p]} \times \mathcal{G}_d^{[q]} \to \mathcal{G}_d^{[p_a^\ell] \cup [p_n^\ell]}$ and $\int [.] ds_{[p_b^\ell]} : \mathcal{G}_d^{[p_a^\ell] \cup [p_n^\ell]} \to \mathcal{G}_d^{[p^\ell]}$, and if the linear component of node p^ℓ is an identity $\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, (.)) \cdot (.) : \underset{p \in par(p^\ell)}{\times} \mathcal{G}^{[p]} \times \mathcal{G}_d^{[q]} \to \mathcal{G}_d^{[p_n^\ell]} := \mathcal{G}_d^{[p^\ell]}$, i.e., such that the application of the derivative of each node preserves the function class.

(iii) For each ℓ, for each p^ℓ, for each input variable $j = 1 \dots 2d_2 + o^\ell$ let $[p_{dj}^\ell] = [p_b^\ell] / ([p_b^\ell] \cap [j])$ indicate the set of integrands integrated over in node p^ℓ excluding the components of input j. For each ℓ, for each p^ℓ, for each input variable $j = 1 \dots 2d_2 + o^\ell$, define recursively along the computational graph the sets $[p_j^{\ell m}]$, $[p_{ej}^{\ell m}]$ for terms $m \in M_j^{p^\ell}$, a set defined as follows. Start at node j with $M_j^j = 1$ and let $[j_j^m] = \emptyset$, $[j_{ej}^m] = \emptyset$. For a node p^ℓ, let $q_j(p^\ell)$ be the set of $q \in par(p^\ell)$ such that q lies along a directed path from node j to node p^ℓ. For each p^ℓ, for each $q \in q_j(p^\ell)$, for each $m' \in M_j^q$, define an $m = m(q, m') \in M_j^{p^\ell}$ where $m(q, m')$ uniquely assigns an index for each parent: e.g. is one to one and onto. If the linear component of p^ℓ is an identity map, $[p_j^{\ell m}] = [p_n^\ell] \cup [q_j^{m'}]$, $[p_{ej}^{\ell m}] = [p_{ej}^{\ell m'}]$. If the linear component of p^ℓ is an integral map, $[p_j^{\ell m}] = ([p_a^\ell] \cup [p_n^\ell] \cup [q_j^{m'}]) / (([p_b^\ell] \cap [q_{ej}^{m'}]) \cup [p_{dj}^\ell])$, $[p_{ej}^{\ell m}] = [p_{ej}^{\ell m'}] \cup ([p_b^\ell] \cap ([j] / [q_{ej}^{m'}]))$.

For all $p^\ell = 1 \dots P^\ell$, $\forall j = 1 \dots 2d_2 + o^\ell$, for all $m \in M_j^{p^\ell}$, let $\hat{\mathcal{G}}_d^{[p_j^{\ell m}];[p_{ej}^{\ell m}]} : \mathcal{S}_{[p_j^{\ell m}]} \times \mathcal{S}_{[p_{ej}^{\ell m}]} \to \mathbb{R}$ be classes of bounded functions containing the constant function 1 and $\forall q \in q_j(p^\ell)$, for all $m' \in M_j^q$ let $\hat{\mathcal{G}}_d^{[p_a^\ell] \cup [p_n^\ell] \cup [q_j^{m'}];[q_{ej}^{m'}]} : \mathcal{S}_{[p_a^\ell] \cup [p_n^\ell] \cup [q_j^{m'}]} \times \mathcal{S}_{[q_{ej}^{m'}]} \to \mathbb{R}$ and $\hat{\mathcal{G}}_d^{[p_n^\ell] \cup [q_j^{m'}];[q_{ej}^{m'}]} : \mathcal{S}_{[p_n^\ell] \cup [q_j^{m'}]} \times \mathcal{S}_{[q_{ej}^{m'}]} \to \mathbb{R}$ be classes of bounded functions such that $k^{p^\ell}(s_{[p_a^\ell]}) \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, (.)) \cdot (.) : \underset{p \in par(p^\ell)}{\times} \hat{\mathcal{G}}^{[p]} \underset{q}{\times} \hat{\mathcal{G}}_d^{[q_j^{m'}];[q_{ej}^{m'}]} \to \hat{\mathcal{G}}_d^{[p_a^\ell] \cup [p_n^\ell] \cup [q_j^{m'}];[q_{ej}^{m'}]}$ if p^ℓ contains an integral operator and $\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, (.)) \cdot (.) :$

$\underset{p\in par(p^\ell)}{\times}\hat{\mathcal{G}}^{[p]} \underset{q}{\times} \hat{\mathcal{G}}_d^{[q_j^{m'}];[q_{ej}^{m'}]} \to \hat{\mathcal{G}}_d^{[p_n^\ell]\cup[q_j^{m'}];[q_{ej}^{m'}]} := \hat{\mathcal{G}}_d^{[p_j^{\ell m}];[p_{ej}^{\ell m}]}$ otherwise.

(iv) For each ℓ, for each p^ℓ, for each $j = 1\dots 2d_2 + o^\ell$, for all $\forall q \in q_j(p^\ell)$, for all $m' \in M_j^q$ let $\hat{\mathcal{G}}_d^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} : \mathcal{S}_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} \to \mathbb{R}$ be classes of bounded functions such that

(a)

$$f(s_{([p_a^\ell]\cup[p_n^\ell]\cup[q_j^{m'}])/(([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell])}, s_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}, s'_{[q_{ej}^{m'}]}) \in \hat{\mathcal{G}}_d^{[p_a^\ell]\cup[p_n^\ell]\cup[q_j^{m'}];[q_{ej}^{m'}]}$$

implies that $\forall s_{([p_a^\ell]\cup[p_n^\ell]\cup[q_j^{m'}])/(([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell])},\ s'_{[q_{ej}^{m'}]}$

$$f(s_{([p_a^\ell]\cup[p_n^\ell]\cup[q_j^{m'}])/(([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell])}, ., s'_{[q_{ej}^{m'}]}) \in \hat{\mathcal{G}}_d^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}$$

(b) the quadrature scheme $\pi^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}$ corresponding to integral $\int[.]ds_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}$ maps $\hat{\mathcal{G}}_d^{[p_a^\ell]\cup[p_n^\ell]\cup[q_j^{m'}];[q_{ej}^{m'}]} \to \hat{\mathcal{G}}_d^{[p_j^{\ell m}];[p_{ej}^{\ell m}]}$.

(c) For each $\ell = 1\dots d_2$, for all $j = 1\dots 2d_2 + o^\ell$, for all $\forall q \in q_j(p^\ell)$, and for all $m' \in M_j^q$ let the quadrature schemes $\pi^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}$ associated with tensor product interpolation schemes $\mathcal{Q}_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}$ be such that for all classes $\hat{\mathcal{G}}_d^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}$

$$\sup_{f\in\hat{\mathcal{G}}_d^{[p_a^\ell]\cup[p_n^\ell]\cup[q_j^{m'}];[q_{ej}^{m'}]}} \sup_{s_{[p_j^{\ell m}];[p_{ej}^{\ell m}]}\in\mathcal{S}_{[p_j^{\ell m}];[p_{ej}^{\ell m}]}}$$

$$\left| \sum_{i=1}^{K_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}} \pi_i^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} f(s_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell],i}, s_{[p_j^{\ell m}];[p_{ej}^{\ell m}]}) \right.$$

$$\left. - \int f(s_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}, s_{[p_j^{\ell m}];[p_{ej}^{\ell m}]}) ds_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} \right| \le \zeta_{dK}^{p^\ell jm}$$

where the rate is $\zeta_{dK}^{p^\ell jm} := \zeta_d^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}(K_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]})$

(v) For each $\ell = 1\dots d_2$, if row ℓ satisfies Condition (4)(b)(ii), for all $j = 1\dots 2d_2$, for all $m \in M_j^{P^\ell}$ let schemes $\mathcal{Q}_{[\chi+1]} \otimes \mathcal{Q}_{[j]}$ be sup norm accurate over classes $\hat{\mathcal{G}}_d^{[P_j^{\ell m}];[P_{ej}^{\ell m}]}$ at rate $\zeta^{\ell jm}(K_{[\chi+1]\times[j]})$

(vi) For each $\ell = 1\dots d_2$, if row ℓ satisfies Condition (4)(b)(i), let $\tilde{n}_{j(out)}^{P^\ell}$ and $n_{j(out)}^{P^\ell}$, defined in Thm (9), be bounded away from 0. For all $j = 1\dots 2d_2$, $m \in M_j^{P^\ell}$, let $\hat{\mathcal{G}}_{dd}^{[P_j^{\ell m}];[P_{ej}^{\ell m}]} : \mathcal{S}_{[P_j^{\ell m}]\underset{m'}{\cup}[P_j^{\ell m'}];[P_{ej}^{\ell m}]\underset{m'}{\cup}[P_{ej}^{\ell m'}]} \to \mathbb{R}$ be classes of bounded functions such that if $f_1 \in \hat{\mathcal{G}}_d^{[P_j^{\ell m}];[P_{ej}^{\ell m}]}$, then $\frac{f_1}{\tilde{n}_{j(out)}^{P^\ell}} \in \hat{\mathcal{G}}_{dd}^{[P_j^{\ell m}];[P_{ej}^{\ell m}]}$. For all $j =$

$1 \ldots 2d_2$ let schemes $\mathcal{Q}_{[\chi+1]} \otimes \mathcal{Q}_{[j]}$ be sup norm accurate over classes $\hat{\mathcal{G}}_{dd}^{[P_j^{\ell m}];[P_{ej}^{\ell m}]}$ at rate $\zeta_d^{\ell j m}(K_{[\chi+1]\times[j]})$.

Remark. Regarding Condition (6), this condition ensures that the functional derivatives are also approximable, and holds under similar conditions to Condition (5), with the differences that I consider different subsets of arguments and that, as the functions of interest are now the derivatives of the structural functions, smoothness conditions on the original function must be stronger to achieve a given rate. For example, if $f(s_1, s_2) \in \Lambda^{\alpha}$, $\alpha > 1$, then $\frac{d}{ds_1} f(s_1, s_2) \in \Lambda^{\alpha-1}$, and if $f(s_{[p^\ell]}) \in \bigotimes_{z \in [p^\ell]} F_{d_z}^{\alpha_z}(\mathcal{S}_z)$, then for any $z_i \in [p^\ell]$, $\frac{d}{ds_{z_i}} f(s_{[p^\ell]}) \in F_{d_{z_i}}^{\alpha_{z_i}-1}(\mathcal{S}_{z_i}) \bigotimes_{z \in [p^\ell]/z_i} F_{d_z}^{\alpha_z}(\mathcal{S}_z)$ and so these function classes suffice. Intuitively, taking derivatives loses up to one derivative of smoothness.

The idea behind conditions (iii)-(v) is that the algorithm distinguishes two types of integrals: integrals taken directly over inputs, for which the operator is represented in spectral form, after interpolation, and integrals taken at intermediate steps, which do not appear in input or output and which are approximated by quadrature. Applying the chain rule repeatedly over the graph, the algorithm constructs for each block (ℓ, j) a representation of the kernel function of the operator, which is a function of inputs $\mathcal{S}_{[j]}$ and outputs $\mathcal{S}_{[\chi+1]}$, at interpolation nodes in these variables, with all intermediate variables in this function integrated out numerically, and uses an interpolation scheme in these variables to construct a spectral representation of the integral operator. The reason to make a distinction between input and intermediate variables is that integrals over input variables must attain uniform accuracy over the full Hilbert space of potential inputs in order to achieve operator norm approximation, while intermediate variables are only integrated at a fixed value, the steady state, for which it is reasonable to assume substantially stronger smoothness assumptions (viz, that these functions are elements of a class $\hat{\mathcal{G}}_d^{([p_b^\ell] \cap [q_{ej}^{m'}]) \cup [p_{dj}^\ell]}$) and so guarantee accuracy of a fixed quadrature scheme.

The reason to distinguish in each term between sets $[p_j^{\ell m}]$ and $[p_{ej}^{\ell m}]$ is that the latter represent those arguments of inputs for which one calculates integrals by quadrature, while the former represent arguments of inputs for which one will compute integrals by interpolation. I must keep track of the difference not only at each node but at each term in the sum of each node because the argument alone may be insufficient to distinguish which units are integrated over by which procedure. The rules for propagating which sets are integrated over and which are interpolated in each component are slightly notationally cumbersome, but at no point in the calculations does one actually need to determine the sets $[p_j^{\ell m}]$, $[p_{ej}^{\ell m}]$, and so on. The integrals represented by quadrature and those represented by spectral methods are determined automatically by the operation of the chain rule over the graph. These subsets and the associated function classes do, however, determine the dimensionality and so the convergence rate of the functions to be approximated and so are useful in the analysis of the procedure and potentially in the choice of representations. The particular accuracy bound

which is needed in (iv)(c) is that the quadrature scheme computes integrals accurately over the intermediate units, uniformly over all other inputs. In the case where the exact quadrature scheme corresponding to the interpolation scheme is used, this is implied by the stronger condition that the interpolation scheme is sup norm accurate over classes $\hat{\mathcal{G}}_d^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}$, because sup norm convergence implies L^1 convergence. However, the reverse is not true, and in some cases the weaker condition is needed to show consistency.

The conditions may be guaranteed in the same way as the corresponding clauses in Condition (5), through the use of multivariate function classes with appropriate properties of composition and marginalization for any relevant subset of arguments. In particular, Hölder smoothness of high enough order for all functions guarantees all conditions for appropriate interpolation schemes including wavelets, splines, and polynomials, and so long as the marginal Hölder condition in (i) is satisfied, extensions can be made to mixed and/or anisotropic classes with corresponding representations.

A simple example may be used to illustrate the distinctions being made between the arguments. Suppose $\mathcal{F}^\ell(g_1(s_1))(t) = \int f^3(t, \int f^2(s_1, g_1(s_1))ds_1, g_1(s_1))ds_1$. In this case, $n^1 = g_1(s_1)$ is the input node, $n^2 = \int f^2(s_1, n^1(s_1))ds_1$ has parent node n^1 and integral linear map, and $n^3(t) = \int f^3(t, n^2, n^1(s_1))ds_1$ has parent nodes n^1 and n^2. In this case, the derivative $\mathcal{F}_g^\ell(g^*(s_1))(t)$ can be computed by the chain rule as

$$\int \frac{d}{dn^2} f^3(t, n^{2*}, n^{1*}(s_1)) \frac{d}{dg_1} n^2[.]ds_1 + \int \frac{d}{dn^1} f^3(t, n^{2*}, n^{1*}(s_1))[.]ds_1 =$$

$$\int \frac{d}{dn^2} f^3(t, n^{2*}, n^{1*}(s_1)) \cdot \int \frac{d}{dn^1} f^2(s_1', n^{1*}(s_1'))[.]ds_1' ds_1 + \int \frac{d}{dn^1} f^3(t, n^{2*}, n^{1*}(s_1'))[.]ds_1'$$

This has an expression as a sum of two integral operators, corresponding to $m = 1, 2 \in M_1^3$ with kernel functions $\int \frac{d}{dn^2} f^3(t, n^{2*}, n^{1*}(s_1)) \cdot \frac{d}{dn^1} f^2(s_1', n^{1*}(s_1'))ds_1$ and $\frac{d}{dn^1} f^3(t, n^{2*}, n^{1*}(s_1'))$ applied to input function $g_1(s_1')$. To represent this numerically, these functions are interpolated and mapped to a set of basis function coefficients. Because the first function contains an integral, this must be approximated as well, which can be performed by numerical quadrature. Note that both kernel functions have arguments (t, s_1'), while the argument s_1 in the first function is integrated out. To see that this corresponds to the above description, note that the arguments in a node p^ℓ which are integrated by quadrature are $([p_b^\ell] \cap [q_{ej}^{m'}]) \cup [p_{dj}^\ell]$. In the case of $p^\ell = 2$, $s_{[p_b^\ell]}$, the integrand, is s_1, which is an input variable so $[p_b^\ell] = [j]$, and so $[p_{dj}^\ell] = \emptyset$, and the single parent satisfies $[q_{ej}^{m'}] = \emptyset$, so the integral is not taken by quadrature. However, now that this integral has been taken, I set argument s_1 to s_1' to indicate that it will be interpolated over and set $[p_{ej}^m] = \{1\}$. In the case of $p^\ell = 3$, $s_{[p_b^\ell]}$, the integrand in node 3, is again s_1, which is an input variable so $[p_b^\ell] = [j]$, and so $[p_{dj}^\ell] = \emptyset$. However, for the first parent node n^2, $[q_{ej}^{m'}] = [1]$ and so $[p_b^\ell] \cap [q_{ej}^{m'}] = [1]$ and the integral over s_1 is calculated by quadrature, while for the second parent node n^1, $[q_{ej}^{m'}] = \emptyset$ and so $[p_b^\ell] \cap [q_{ej}^{m'}] = \emptyset$ and so the integral over s_1 is not calculated

by quadrature and is instead interpolated over.

The following lemma shows that the above conditions ensure that a reasonable approximation of the steady state also ensures a reasonable approximation of all node values at that point. Two approximations of the exact steady state value of the node function n^{*p^ℓ} are used: the numerically feasible discrete approximation $\overrightarrow{n}^{*p^\ell}$ produced by the algorithm up to node p^ℓ for the value of the node on the grid and the intermediate construction $\tilde{n}^{p^\ell}$ derived for analytical convenience consisting of the function which would be produced using interpolated approximate inputs $\tilde{g}^*_j$ and applying the node functions up to node p^ℓ, replacing the exact integrals with quadrature. This latter object need not be computed in reality, but provides a book-keeping tool to keep track of the numerical error introduced and propagated through each step of the algorithm.

Lemma 7. *Under Conditions (3), (4), and (5) (i)-(iii), defining* $\forall \ell,\ \forall p^\ell = 1 \dots 2d_2 + o_\ell$, $n^{*p^\ell} = g^*_{j(p^\ell)}$, $\overrightarrow{n}^{*p^\ell} = \overrightarrow{g}^*_{j(p^\ell)}$, *and* $\tilde{n}^{*p^\ell} = \tilde{g}^*_{j(p^\ell)}$, *and for* $p^\ell > 2d_2 + o_\ell$ *defining* n^{*p^ℓ} *and* $\overrightarrow{n}^{*p^\ell}$ *inductively along the computational graph as* $n^{p^\ell}(\{n^{*p}\}_{p \in par(p^\ell)})$, $\overrightarrow{n}^{p^\ell}(\{\overrightarrow{n}^{*p}\}_{p \in par(p^\ell)})$, *respectively, and* $\tilde{n}^{*p^\ell}$ *as*

$$f^{p^\ell}(s_{[p^\ell_I]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p \in par(p^\ell)})$$

when the linear component corresponding to p^ℓ *is* I, *and*

$$\sum_{i=1}^{K_{[p^\ell_b]}} \pi_i^{[p^\ell_b]} k^{p^\ell}(s_{[p^\ell_a],i}) f^{p^\ell}(s_{[p^\ell_I],i}, \{\tilde{n}^{*p}(s_{[p],i})\}_{p \in par(p^\ell)})$$

when the component is an integral operator, then

(i) $\forall \ell,\ \forall p^\ell\ n^{*p^\ell} \in \mathcal{G}^{[p^\ell]}$.

(ii) Assume also (5) (iv)-(vi). For any ℓ, $\forall p^\ell$, *define the variable* $1_{p^\ell} := 0$ *if the node* $p^\ell > 2d_2 + o^\ell$ *and has the identity map for linear component, and 1 otherwise, let* $dp(p^\ell)$ *be the set of directed paths terminating in node* p^ℓ, *and* $\forall j \in dp(p^\ell)$, $\forall p \in j$, *let* $desc(p, j)$ *be the set of edges* (a, b) *from a node* a *to a node* b *downstream of node* p *in path* j, *and define* $\alpha^{p,j} = \prod_{(a,b) \in desc(p,j)} \alpha^{b(a)}$. *For convenience, abuse notation to define* $\zeta^{p^\ell}(K_{p^\ell}) = \zeta^{[p^\ell_b]}(K_{[p^\ell_b]})$ *if* $p^\ell > 2d_2 + o^\ell$ *(and as defined in (5) (vi) otherwise). It is the case that*

$$\left\| \tilde{n}^{*p^\ell} - n^{*p^\ell} \right\|_\infty \leq \upsilon^{p^\ell} := C(\zeta^{p^\ell}(K_{p^\ell}) 1_{p^\ell} + \sum_{j \in dp(p^\ell)} \sum_{p \in j} (\zeta^{p(e)}(K_{p(e)}) 1_p)^{\alpha^{p,j}}) \quad (27)$$

$$\left\| \sum_{k=1}^{K_{[p^\ell]}} (M_{[p^\ell]} \overrightarrow{n}^{*p^\ell})_k \phi_k - n^{*p^\ell} \right\|_\infty \leq C(\zeta^{[p^\ell]}(K_{[p^\ell]}) + (\zeta^{p^\ell}(K_{p^\ell}) 1_{p^\ell} + \sum_{j \in dp(p^\ell)} \sum_{p \in j} (\zeta^{p(e)}(K_{p(e)}) 1_p)^{\alpha^{p,j}})) \quad (28)$$

where C are constants not depending on K and $p(e)$ indexes the node corresponding to the origin of directed edge e and $p(e)$ indexes the set of arguments to be integrated over in that node.

Remark 8. Note that because the computational graph has finite cardinality, the number of directed paths terminating in any given node is finite, so these sums are also bounded. The constant term in each inequality depends on the node, and in particular on the set of directed paths leading to it and their regularity. In practice, the constant in the latter inequality will be larger than that in the former, though I do not make this explicit. In general, the bound will be driven by the worst case rate $(\zeta^{p_n(e)}(K_{[p_n(e)]}))^{\alpha_{p,j}}$ among any set of edges.

Given convergence of the approximate steady state values around which the operators are differentiated, convergence of the functional derivatives and, in consequence, of the approximate solutions follows from an analogous inductive argument under similar conditions, viz Condition (6). The following lemma demonstrates that Algorithm (1) converges, and defines the convergence rates in terms of model primitives.

Lemma 9. *Let the model of interest satisfy conditions (3), (4), (5), and (6)(i)-(iv). Defining $\forall \ell$, $\forall p^\ell = 1 \dots 2d_2 + o_\ell$, for all $j = 1 \dots 2d_2 + o^\ell$, $n_j^{p^\ell 1} = 1\{p^\ell = j\}$, and $\tilde{n}_j^{p^\ell 1} = 1\{p^\ell = j\}$, and for $p^\ell > 2d_2 + o_\ell$ defining $n_j^{p^\ell m}$ inductively along the computational graph for all $q \in q_j(p^\ell)$, for all $m' \in M_j^q$ as $n_j^{p^\ell m} := \int k^{p^\ell}(s_{[p_a^\ell]}) \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \cdot n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) ds_{([p_b^\ell] \cap [q_{ej}^{m'}]) \cup [p_{dj}^\ell]}$ when the linear component corresponding to p^ℓ is an integral operator, and $n_j^{p^\ell m} := \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \cdot n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]})$ when the component is I and $n_j^{p^\ell} = \sum_{m \in M_j^{p^\ell}} n_j^{p^\ell m}$. Similarly, for $p^\ell > 2d_2 + o_\ell$, for all $q \in q_j(p^\ell)$, for all $m' \in M_j^q$ define inductively $\tilde{n}_j^{p^\ell m}$ for $m = m(q, m')$ as*

$$\sum_{i=1}^{K_{([p_b^\ell] \cap [q_{ej}^{m'}]) \cup [p_{dj}^\ell]}} \pi_i^{([p_b^\ell] \cap [q_{ej}^{m'}]) \cup [p_{dj}^\ell]} k^{p^\ell}(s_{[p_a^\ell],i})$$
$$\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell],i}, \{\tilde{n}^{*p}(s_{[p],i})\}_{p \in par(p^\ell)}) \cdot \tilde{n}_j^{qm'}(s_{[q_j^{m'}]i}, s'_{[q_{ej}^{m'}]})$$

when the linear component corresponding to p^ℓ is an integral operator, and

$$\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \cdot \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]})$$

when the component is I. As before, define $\tilde{n}_j^{p^\ell} = \sum_{m \in M_j^{p^\ell}} \tilde{n}_j^{p^\ell m}$

For any ℓ, $\forall p^\ell$, define the variable $1_{p^\ell} := 0$ if the node $p^\ell > 2d_2 + o^\ell$ and has the identity map for linear component, and 1 otherwise. Let $dp(j, p^\ell, m)$ be the multiset of pairs (p, m') corresponding to terms $n_j^{pm'}$ along any directed

path originating in node j and terminating in term $n_j^{p^\ell m}$ defined by iterating the (unique) inverse map of $m(q,m')$ back to node j. It is the case that $\forall \ell$, $\forall p^\ell$, $\forall j = 1 \dots 2d_2 + o^\ell$, $\forall m \in M_j^{p^\ell}$

$$\left\| \tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')} \right\|_\infty \leq C \sum_{(p,\tilde{m}) \in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}})$$

implying that

$$\left\| \tilde{n}_j^{p^\ell} - n_j^{p^\ell} \right\|_\infty \leq C \sum_{m \in M_j^{p^\ell}} \sum_{(p,\tilde{m}) \in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}}) \quad (29)$$

where C is a constant not depending on K.

Assume also conditions (6)(v)-(vi). $\forall \ell = 1 \dots d_2$, if ℓ satisfies Condition (4)(b)(ii), $\forall j = 1 \dots 2d_2$

$$\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} (M_{[\ell^o]} \mathcal{F}^\ell_{\overrightarrow{g}_{j(in)}} (\Pi^{[j]})^{-1} M^*_{[j]})_{ik} \phi_i \phi_k - n_j^{P^\ell} \right\|_\infty \leq \epsilon_K^{(\ell,j)}$$

$$:= C \sum_{m \in M_j^{P^\ell}} (\zeta^{\ell jm}(K_{[\chi+1]\times[j]}) + \sum_{(p,\tilde{m}) \in dp(j,P^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}})) \quad (30)$$

and if ℓ satisfies Condition (4)(b)(i), $\forall j = 1 \dots 2d_2$

$$\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} (M_{[\ell^o]} (\mathcal{F}^\ell_{\overrightarrow{g}_{j(p^\ell(out))}})^{-1} \mathcal{F}^\ell_{\overrightarrow{g}_{j(in)}} (\Pi^{[j]})^{-1} M^*_{[j]})_{ik} \phi_i \phi_k - n_j^{P^\ell} / n_{j(out)}^{P^\ell} \right\|_\infty \leq \epsilon_K^{(\ell,j)} :=$$

$$C(\sum_{m \in M_j^{P^\ell}} \zeta_d^{\ell jm}(K_{[\chi+1]\times[j]}) + \sum_{m \in M_j^{P^\ell}} \sum_{(p,\tilde{m}) \in dp(j,P^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}})$$

$$+ \sum_{m \in M_{j(out)}^{P^\ell}} \sum_{(p,\tilde{m}) \in dp(j(out),P^\ell,m)} (\zeta_{dK}^{pj(out)\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}})) \quad (31)$$

where C are constants not depending on K.

The above result ensures that the approximated functional derivatives of an equilibrium system constructed by Algorithm (1) are such that any component of the functional derivative which takes the form of a kernel integral operator is represented by an approximate kernel function which converges uniformly. The setup of the algorithm also ensures that any component of the functional derivative which takes the form of an identity operator is represented as an identity matrix. An application of Young's inequality leads to the immediate corollary that the derivative approximation converges in operator norm at the

same rate. This is a key condition needed for the rational expectations solution of the linearized system to likewise converge in operator norm, supposing that conditions described in Childers [2018] which ensure global existence, uniqueness, and continuity of this solution also hold.

Theorem 10. *Assume the conditions of Lemma (9). In addition assume the full system satisfies Condition (2) in Childers [2018]. Then Algorithm (1) produces approximate policy operators* $\tilde{h}_x^K$, $\tilde{g}_x^K$ *which satisfy the conditions of Childers [2018] Theorem 1 and so satisfies, for* $\mathcal{H}^K = \underset{j \in 1 \ldots d_x}{\times} Sp \otimes_{z \in [j]} \Phi_z$, *and* $\epsilon_K = \max_{\ell \in 1 \ldots d_2, j = 1 \ldots 2d_2} \epsilon_K^{(\ell,j)}$, $\exists \bar{K}$ *such that for all* $K \geq \bar{K}$

$$\sup_{\|f\|_{\mathcal{H}^K}=1} \left\| (\tilde{h}_x^K - h_x)(f) \right\| \leq O(\epsilon_K)$$

$$\sup_{\|f\|_{\mathcal{H}^K}=1} \left\| (\tilde{g}_x^K - g_x)(f) \right\| \leq O(\epsilon_K)$$

Remark 11. The constant in these bounds is in principle dependent on the number of variables d_2, with worst case dependence $O(d_2^2 \epsilon_K)$, though this may be pessimistic in the case of block-wise sparsity. The restriction to space $\mathcal{H}^K$ arises because the algorithm computes only components in the span of the basis representation. One may construct an approximation to the solution operators on the orthogonal span of this space by the analytical procedure described in Childers [2018] if desired, though this does not change predictions for any inputs in space $\mathcal{H}^K$. Regarding Condition (2) in Childers [2018], asymptotic diagonality is fulfilled by any model satisfying the conditions of Lemma (9). The other conditions, which rule out roots of unity and ensure existence, uniqueness, and continuity of local solutions, are global properties of the entire system and may not be verified solely from the properties of the constituent functions. These conditions constitute generalizations of the eigenvalue conditions of Blanchard and Kahn [1980], and so may be violated in models which do not satisfy stability or uniqueness properties, for example, monetary models under passive policy. See Childers [2018] for a discussion of sufficient conditions and potential for verifiability of these requirements.

4 Examples and Results

The algorithm and guarantees in the previous section present general purpose methodology for setting up and solving a rational expectations model with variables which may be functions but in particular applications, the conditions must be verified and decisions made regarding choice of tuning parameters including choice of interpolation method and number K of grid points to use in each dimension. In this section, I move back to the specialized but still broadly applicable setup of Section (2.1) to describe how the algorithm can be implemented and conditions verified in models of this more limited class, and provide an illustration via the Huggett-style model of Section (2.2). In any particular case,

the properties of the algorithm may depend on details of the setup and even particular parameter values at which the model is evaluated. For example, the conditions of Theorem (10) which ensure that an accurate approximation of the linearized first order conditions also ensures an accurate solution of the model as a whole depend on conditions analogous to the Blanchard and Kahn [1980] eigenvalue conditions for existence and uniqueness of a locally stable solution, which may depend on parameters of both exogenous and endogenous processes.

Conditional on a model with solution which is well defined and stable, the properties of the method depend primarily on the suitability of the function approximation method used for representing the functions which constitute the model. Appendix (A) collects a large set of pairs of function classes and interpolation methods and provides results which can be used to bound the ζ and ζ_d terms in the bounds from Lemmas (7) and (9) which determine this rate. These are mainly collected from the literature on function approximation but with a few new results. These are summarized in Table (1), with precise definitions of each method and class and statement of convergence theorems provided in Appendix (A).

A summary of the theoretical and numerical results over a selection of models is that in cases where these functions are highly regular, the algorithm appears to exhibit strong performance with only moderately fine discretization, in both theory and practice. Tensor product Chebyshev polynomial and spline methods perform well in cases of moderate regularity, and Chebyshev methods perform especially well in cases of high regularity such as analytic functions. In high dimensions, additional gains can be obtained through the use of sparse grid methods. In cases where the functions are irregular, much finer discretization appear to be necessary for reasonable performance. This is reflected in the theory for nonsmooth or discontinuous functions, in which cases local methods such as histograms or wavelets may be required or at least superior to smooth methods like polynomials, which can exhibit unavoidable approximation errors, as demonstrated by the well known Gibbs phenomenon which leads to approximation artifacts in discontinuous functions. In cases in which the functions possess many derivatives but high curvature and narrow support, performance also appears poor until a very fine discretization is used. This would likely be reflected more precisely by quantifying the constant terms in front of the rates of convergence, suggesting that the rates of convergence should be seen to reflect an asymptotic regime, which in some cases may occur at a degree of approximation which is beyond that which is computationally practical in some applications. Whether this reflects an impassible difficulty specific to particular models or could be resolved using alternative methods appears to be open. While lower bounds for the constituent function approximation problems are often available which suggest that each step may be near optimal, for procedures such as this which compose many applications of function approximation, it may be the case that not all steps need to be performed as described to obtain a valid final result.

The case in which some of the functions contain discontinuities, including in the Huggett model of section (2.2) merits additional caution. In many such cases, no feasible sup norm consistent interpolation method may exist, and so

a modified version of Lemmas (7) and (9) must be used instead, with bounds identical in form but with convergence in slightly different (weaker) norms. Two approaches are provided, both based on histogram representations: one which only works with a limited class of discontinuous functions but results in solution operators which converge in the same sense as in Theorem (10), and another which allows a broader set of possible discontinuities, but provides weaker guarantees for the solution operators. In both cases, the procedure actually implemented is exactly that of Algorithm (1) using evenly spaced histograms as the interpolation approach: it is only the analysis and guarantees which differ. The latter approach is the one which must be used in the model of Section (2.2), theoretical and numerical results for which are given in the section (4.2).

4.1 Application to Canonical Model

While the algorithm and results provided were designed with the general structure of the canonical class of heterogeneous agent models of Section (2.1) in mind, many features of the method extend beyond what is necessary in order to accommodate basic examples from this class. In basic versions of this model, only very shallow paths with no more than two or three iterated nodes are required, with few intersecting structures. Simple modifications can feasibly lead to arbitrary depth or complexity. For example, utility functions can contain a neural network of arbitrary depth over some possibly infinite dimensional input; nested CES and nested logit preferences (using an ex ante value function) present simple special cases of this construction. Neural network constructions can similarly be used inside transition function Q to represent a variety of dynamic patterns, so long as invertibility is maintained, or inside market clearing conditions or dynamics of exogenous variables. The feasibility of such extensions follows straightforwardly from the conditions of the method and may lead to novel model structures, or at least additional flexibility within existing models.

Conversely, with the standard class, some implementation details must be specialized in order to ensure validity of the procedure. In particular, the form of the equilibrium conditions, which result in a linearized system with formulas expressed in Appendix (B.1), imposes certain choices regarding the representations chosen, the upshot of which is that the interpolation matrices $M_{[j]}$ must be chosen in a particular way. The variables, $f_X(.)$ and $Lf_X(.)$, which represent the cross-sectional density of idiosyncratic state X_i can be represented using whichever interpolation scheme is appropriate to the smoothness properties of the model, but it is important that normalization is imposed so that perturbations of densities integrate to 0, as implemented, e.g., by Algorithm (3). In addition to aiding interpretation of perturbations as densities, this removes a redundant equation from the linearized system. As any correctly specified Kolmogorov Forward equation will ensure densities remain normalized, the action on the basis function corresponding to a constant function will be to keep it constant. This results in a spurious generalized eigenvalue equal to 1 (or, depending on the approximation used, approximately equal to 1), which may be recognized by a linear rational expectations solver as a violation of the required

Table 1: Basis Functions, Function Spaces, and Convergence Rates

Interpolation Method	Function Class	Convergence Rate	Notes
Univariate Chebyshev Polynomials	$\{f\ [-1,1] \to \mathbb{R} : f^{(v)}$ absolutely continuous $\forall v < \nu,\ \int \left\lvert f^{(\nu)}(s)\right\rvert ds \leq V\}$	$\frac{4V}{\pi v(K+1-v)^v}$	
	$\{f\ [-1,1] \to \mathbb{R} :\ f$ has analytic continuation over complex ellipse of radius $\rho\}$	$\frac{4M\rho^{-(K+1)}}{\rho-1}$	
Tensor Product Chebyshev	$\mathcal{F}^\alpha = \{f\ [-1,1]^d \to \mathbb{R} :\ \omega_z(t) := \sup_{\lvert s_z - s_z^*\rvert \leq t} \lvert f(s_1,\ldots,s_z,\ldots s_d) - f(s_1,\ldots,s_z^*,\ldots s_d)\rvert \leq c_z t^{-\alpha_z}$ for some $\alpha_z > 0$, $c_z < \infty\ \}$	$O((\frac{2}{\pi})^d \prod_{z=1}^d \log K_z \cdot \sum_{z=1}^d \omega_z(\frac{1}{K_z+2}))$	
Histograms	$\Lambda^\alpha([0,1)^d)$ d-dimensional $\alpha-$Hölder	$O(d^{\frac{\min\{\alpha,1\}}{2}} K^{\frac{-\min\{\alpha,1\}}{d}})$	
	$\mathcal{PR}_{1,0}^{\alpha,N}([0,1)^{d_x},[0,1)^{d_y})$ / $\mathcal{PR}_{0,1}^{\alpha,N}([0,1)^{d_x},[0,1)^{d_y})$	$O(K^{-\frac{\min\{\alpha,1\}}{d_x+d_y}})$	Rate is in $\lVert.\rVert_{\infty/1}/\lVert.\rVert_{1/\infty}$ norm only: see section (A.3)
Histograms (transformed)	$\mathcal{PR}_{0,0}^{\alpha,N}([0,1)^{d_x},[0,1)^{d_y})$	$O(K^{-\frac{\min\{\alpha,\frac{1}{2}\}}{d_x+d_y}})$	Infeasible: requires additional step: see section (A.3)
Coiflet or Daubechies Wavelets	$\Lambda_{per}^\alpha([0,1)^d)$ periodic d-dimensional $\alpha-$Hölder	$O(K^{\frac{-\min\{\alpha,s\}}{d}})$	$s = m$ for coif(m), $s = \min\{m,2\}$ for db(m)
Tensor Product Splines	$X_0^{\infty,p+1}([0,1]^d) = \{f \in C^{p+1}([0,1]^d)$, compact support, $\lVert D^\alpha u\rVert_\infty < \infty\ \forall\alpha$ s.t. $\lvert\alpha\rvert_\infty \leq p+1\}$	$O(d(\frac{0.5585}{2^{p+1}-1})^d K^{-\frac{p+1}{d}})$	
Tensor Product Smolyak Grid Splines	$X_0^{\infty,p+1}([0,1]^{\frac{d}{2}} \times [0,1]^{\frac{d}{2}})$	$O(K^{-(p+1)/2} \lvert\log_2 K\rvert^{(p+2)(\frac{d}{2}-1)})$	Symmetric tensor product of 2 Smolyak grids

eigenvalue conditions for uniqueness. As this variation has no physical meaning, it can and should be removed by normalization.

Another issue which may sometimes occur when implementing a heterogeneous agents model which may require attention is the presence of "redundant" variables, and particularly, the use of both a value function and a policy function as state variables when the decision rule may be described in full by either one individually. This creates multiple issues. The simplest, and least worrisome, is that if one variable can be solved out of the system as a function of another, the size of the derivative matrices might be reduced, improving the speed of the algorithm. This provides sufficient justification for reducing the Bellman equation and first order conditions to just an Euler equation, or, in those cases where the policy can be solved for explicitly, just a Bellman equation. The second and more serious issue is that the presence of both equations can cause violation of Condition (4)(b) which restricts the classes of maps present in each equilibrium condition. The issue is that the Bellman equation (6) has derivatives with respect to both V_t and $g_t(X, \epsilon)$ which are not compact operators, and so one can not divide through to produce an equation defined only in terms of identity and kernel integral operators. A straightforward resolution of this is to solve out explicitly to remove one of these as a state variable, and define the other as a function of this chosen state. An alternative resolution, which has the advantages of requiring no additional symbolic calculations and being applicable in models where no closed form representation is even feasible, is to solve implicitly. Luckily, this can be done using a standard application of Algorithm (1), with a particular choice of interpolation matrix M for variable $g(X, \epsilon)$. A similar issue can arise also within the first order conditions determining $g(X, \epsilon)$ when some choice variables Y_{it} correspond to "static" intratemporal decisions which are not directly influenced by expected future payoffs. These can often be solved out of the system as explicit functions of current period variables, but even when this is not feasible, can also be solved out implicitly in standard application of Algorithm (1).

The idea behind the approach is that so long as the conditions of the algorithm apply to a version of the model in which the function $g_t(X, \epsilon)$ has been solved out, any procedure which results in a numerically equivalent answer will inherit the same guarantees. One can then exploit the fact that, because one seeks only a first order solution, even when no closed form solution exists to solve out a redundant variable, one only needs to find the first order Taylor expansion of this formula, which can be calculated by the implicit function theorem applied to the first order conditions. As a result, one can replace $g_t(X, \epsilon)$ by its Taylor expansion in any place it occurs in the other equilibrium conditions before linearizing and obtain a set of derivatives identical to those obtained by using an explicit solution rather than just the first derivative thereof. This approach can be implemented explicitly by first linearizing the first order condition equation (7) and solving for $g_t(X, \epsilon)$, then replacing $g_t(X, \epsilon)$ in all the other equations with this linearized solution.[10] This requires an additional analytical

[10]The author thanks Keshav Dogra for suggesting this approach in a related context.

step, but removes a variable and an equation from the system and so might improve computational speed. Alternately, one can achieve numerically equivalent results without adding a step to the algorithm, by leaving $g_t(X, \epsilon)$ in as a state variable, in which case the substitutions are performed automatically. The one issue which must be addressed is ensuring that the post processing which maps function values at grid points to coefficients properly handles these substitutions. Under reasonable conditions, this can be achieved by using for this variable the same grid points as used for the approximation of the value function $V_t(X, \epsilon)$, but using an identity matrix in place of the interpolation matrix M and the matrix Π^{-1} in parts of the algorithm where these would be applied to the function $g_t(X, \epsilon)$. This ensures that the resulting solution is the same as the one that would be obtained by solving out explicitly for $g_t(X, \epsilon)$, with no unnecessary applications of an interpolation matrix. The resulting solution will express $g_t(X, \epsilon)$ as a function of predetermined variables, expressed as a set of grid points. To recover the coefficients, one can multiply ex post by the interpolation matrix M used for the value function.

The following set of conditions describe a set of sufficient conditions under which the application of Algorithm (1) to the model of Section (2.1), with particular choice of interpolation schemes, results in a consistent procedure.

Condition 12. (i) Let equilibrium conditions be defined by equations (6), (7), (10), (11), (13), and (15) with predetermined state variables $Lf_{X,t}(X_{it})$, $Lg_t(X_{it}, \epsilon_{it})$, LP_t, P_{2t} in $\mathcal{H}_x = L^2_0(\mathcal{X}) \times L^2(\mathcal{X} \times \epsilon) \times \mathcal{H}_P \times \mathcal{H}_{P_2}$ and jump variables $f_{X,t}(X_{it})$, $g_t(X_{it}, \epsilon_{it})$, P_{1t}, and $V_t(X_{it}, \epsilon_{it})$ in $\mathcal{H}_y = L^2_0(\mathcal{X}) \times L^2(\mathcal{X} \times \epsilon) \times \mathcal{H}_{P_1} \times L^2(\mathcal{X} \times \epsilon)$, where $\mathcal{X} \times \epsilon$ are bounded subsets of $\mathbb{R}^{n_x}, \mathbb{R}^{n_\epsilon}$. Let $u(Y_i, X_i, P, \epsilon_i)$ be twice continuously differentiable with 1-Hölder second derivatives, and strictly concave in Y_i. Let the functional derivative of first order condition equation (7) possess bounded inverse for all (X_{it}, ϵ_{it}).[11] Let $(P_1, P_2) \in \mathcal{H}_{P_1} \times \mathcal{H}_{P_2}$ where each space is a Cartesian product of Euclidean space and or functions over (spaces isomorphic to) $\mathcal{X}$ or ϵ. Let $u(.)$, $f_\epsilon(.)$, and $Q(.)$ depend either only on a finite dimensional subvector of P_t and/or P_{t+1}, or, if P_t is function valued, let all paths through these functions contain an integral operator with bounded 1-Hölder kernel function. Let f_U, $\frac{d}{dU} f_U$, f_ϵ, and $Q^{-1}_{y,x}(.)$, and their derivatives and the determinants thereof be 1-Hölder in their arguments. Let market clearing function $F(f_{X,t}(X_{it}), g_t(X_{it}, \epsilon_{it}), P_t)$ on its own satisfy Condition (4)(b)[12] and be composed of functions which are $1-$Hölder in all arguments.

(ii) Let $(\mathcal{T}^{\mathcal{X}}_{K_x}, M^{\mathcal{X}}_{K_x}, \Phi_{K_x}(X))$ be a linear order K_x interpolation scheme over $\mathcal{X}$ sup norm accurate at rate ζ_{K_x} over $1-$Hölder functions on $\mathcal{X}$. Let $(\mathcal{T}^{\epsilon}_{K_\epsilon}, M^{\epsilon}_{K_\epsilon}, \Phi_{K_\epsilon}(\epsilon))$ be an order K_ϵ interpolation scheme over ϵ sup norm accurate at rate ζ_{K_ϵ} over $1-$Hölder functions on ϵ, and let the tensor product $(\mathcal{T}^{\mathcal{X}}_{K_x} \otimes \mathcal{T}^{\epsilon}_{K_\epsilon}, M^{\mathcal{X}}_{K_x} \otimes M^{\epsilon}_{K_\epsilon}, \Phi_{K_x}(X) \otimes \Phi_{K_\epsilon}(\epsilon))$ be an order $K_\epsilon \times K_x$ interpolation scheme over $\mathcal{X} \times \epsilon$ which is sup norm accurate at rate $\zeta_{K_\epsilon \times K_x}$ over $1-$Hölder

[11] See appendix (B.1) for formula: this derivative is a multiplication operator, so this is equivalent to assuming the function is bounded away from 0.

[12] This can be weakened to the condition that after replacing $g_t(X, \epsilon)$ by the solution of first order condition equation (7), F satisfies this condition.

functions on $\mathcal{X} \times \epsilon$. Let the tensor product of this scheme be sup norm accurate at rate $\zeta_{K_\epsilon^2 \times K_x^2}$. Let the corresponding interpolation schemes over $\mathcal{H}_{P_1}$ and $\mathcal{H}_{P_2}$ be of dimensions K_{P_1} and K_{P_2} respectively, and let them and tensor products thereof with the previous schemes be sup norm consistent over 1-Hölder functions over the corresponding spaces at rates $\zeta_{K_{P_1} \times K_x \times K_\epsilon}$, $\zeta_{K_{P_2} \times K_x \times K_\epsilon}$, $\zeta_{K_{P_1} \times K_{P_1}}$, $\zeta_{K_{P_2} \times K_{P_2}}$, $\zeta_{K_{P_1} \times K_x}$, $\zeta_{K_{P_2} \times K_x}$. Denote the maximum of these rates as ζ_P. Let in all cases the quadrature scheme used be the exact quadrature scheme associated with the interpolation scheme.

(iii) Let a steady state of the model exist and be 1-Hölder, with steady state values of functions in x, y at grid points of the schemes from condition (ii) calculated by any procedure, with sup norm error ζ_K.

The above conditions are fairly mild, but do impose a few restrictions. Smoothness conditions rule out models with kinks or discontinuities, including the Huggett model example. Stronger conditions than these could be imposed to achieve correspondingly faster rates; the choice of Hölder conditions is mainly for simplicity, as it is the simplest condition under which one need not consider the topology of the computational graph when calculating the rate of convergence, and also ensures that all of the various function classes described in Conditions (5) and (6) can be taken to be the class of Hölder functions, which satisfies all the required conditions. Strong concavity and nonnegativity conditions ensure that first order conditions are necessary and sufficient for unique interior solutions; relaxing these conditions may introduce multiplicity or corner solutions, which must be dealt with by other methods. The conditions on the market clearing conditions are left abstract as these may take many forms. In cases in which these variables are finite dimensional and affected only by weighted averages of individual states or actions, as is typical of general equilibrium or mean field interactions, these conditions are always satisfied. In cases where the aggregate states are themselves functions, as in general equilibrium models with a continuum of goods (for example, trade models where goods are indexed by location), these conditions might in principle rule out certain forms of purely local interactions, but appear to be satisfied by a broad variety of standard models.

The condition on the interpolation schemes is weak, satisfied by a broad variety of tensor product schemes including splines, wavelets, Chebyshev polynomials, histograms, and so on. The steady state condition does not appear particularly onerous so long as such a state exists, noting that fixed points generally preserve the Hölder property, but verifying a rate of convergence may depend on fine details of the model and steady state solution algorithm which are left unspecified here and so this is left as a high level condition. In practice, a procedure such as the iterative algorithm introduced in Huggett [1993] is recommended.

Theorem 13. *Let a model satisfy condition (12)(i). Apply Algorithm (1) to this model, using interpolation schemes and the corresponding exact quadrature weights satisfying condition (12)(ii) over all variables with the following changes:*

In step (5), $M_{[j]} = M^{\mathcal{X}}_{K_x} \otimes M^{\epsilon}_{K_\epsilon}$ should be replaced by $I_{K_\epsilon \times K_x}$ and $\Pi^{[j]}$ by $I_{K_\epsilon \times K_x}$ in the columns j corresponding to $g(X,\epsilon)$ or $Lg(X,\epsilon)$, and $M_{[\ell^o]} = M^{\mathcal{X}}_{K_x} \otimes M^{\epsilon}_{K_\epsilon}$ replaced by $I_{K_\epsilon \times K_x}$ in the first order condition equations (7). $g_t(X_{it}, \epsilon_{it})$ should be counted as an input argument in Bellman equation (6) (explicitly: it should be premultiplied by $M_{[\ell^o]} = M^{\mathcal{X}}_{K_x} \otimes M^{\epsilon}_{K_\epsilon}$).[13]

Let $K_{[x]} = K_x - 1 + K_x \times K_\epsilon + K_{P_1} + 2K_{P_2}$, $K_{[y]} = K_x - 1 + K_x \times K_\epsilon + K_{P_1} + K_x \times K_\epsilon$ be the number of grid points used for predetermined and jump variables, respectively.

Define $V_{[x]} \in \mathbb{R}^{K_{[x]} \times K_{[x]}}$ as the block diagonal matrix with $M^{\mathcal{X}}_{K_x} \otimes M^{\epsilon}_{K_\epsilon}$ in the $K_x : K_x + K_x \times K_\epsilon$ entries (corresponding to Lg) and identities in other blocks, and $V_{[y]} \in \mathbb{R}^{K_{[y]} \times K_{[y]}}$ as the block diagonal matrix with $M^{\mathcal{X}}_{K_x} \otimes M^{\epsilon}_{K_\epsilon}$ in the $K_x : K_x + K_x \times K_\epsilon$ entries (corresponding to g) and identities in other blocks.

In Step (6), pre- and post- multiply the matrices h^K_x, g^K_x with $V_{[x]} h^K_x V^*_{[x]}$ and $V_{[y]} g^K_x V^*_{[x]}$ before constructing maps $\tilde{h}^K_x$, $\tilde{g}^K_x$.

In addition assume the full system satisfies Condition (2) in Childers [2018]. Then this modified version of Algorithm (1) produces approximate policy operators $\tilde{h}^K_x$, $\tilde{g}^K_x$ which satisfy, for $\mathcal{H}^{K_{[x]}} = Sp\ \Phi_{K_x} \times Sp\ (\Phi_{K_x} \otimes \Phi_{K_\epsilon}) \times Sp\ \Phi_{K_{P_1}} \times Sp\ \Phi_{K_{P_2}} \times Sp\ \Phi_{K_{P_2}}$, $\exists \bar{K}$ such that for all $K \geq \bar{K}$

$$\sup_{\|f\|_{\mathcal{H}^K}=1} \left\| (\tilde{h}^K_x - h_x) f \right\| \leq O(\max\{\zeta_{K^2_\epsilon \times K^2_x}, \zeta_K, \zeta_P\})$$

$$\sup_{\|f\|_{\mathcal{H}^K}=1} \left\| (\tilde{g}^K_x - g_x) f \right\| \leq O(\max\{\zeta_{K^2_\epsilon \times K^2_x}, \zeta_K, \zeta_P\})$$

Remark. This result says that so long as the variable $g(X,\epsilon)$ is represented using grid points rather then function interpolation, one can apply the baseline solution method, ignoring the fact that Condition (4)(b) fails in the setup of the model, because the results are numerically equivalent to those that would be obtained by solving to remove $g(X,\epsilon)$ and $Lg(X,\epsilon)$ from the system. The only change to be made is that, because these variables are represented by grid points, to recover their behavior as well, they must be transformed to coefficients by interpolation, which will recover the relationship that could have been found by applying the implicit function theorem to the first order conditions. If behavior of $g_t(X,\epsilon)$ is not desired, the modification of step (6) is not needed; one could simply remove the associated rows and columns from the solutions and use the standard representation for the remaining variables.

4.2 Huggett Model Theory and Results

The consumption savings model from Section (2.2) while a special case of the general class, presents some additional challenges. The most important of these

[13] In case market clearing equation (10) satisfies the weaker condition of footnote (12), if there exists a sub equation in which the derivative with respect to $g(X,\epsilon)$ is a multiplication operator, it should also be treated as an input argument.

is the presence of a kink in, and so nondifferentiability in the derivative of, the consumption policy rule as a function of wealth at a point which is determined in equilibrium, which prevents the possibility of use of an interpolation method which converges uniformly over its domain. This does not prevent the use of Algorithm (1) to construct a valid solution, but does require a slightly modified analysis, which produces weaker guarantees. So long as one is willing to replace an approximation guarantee which holds uniformly over all square integrable functions with one which holds only over Hölder continuous functions, certain classes of interpolation method can continue to produce consistent results. In particular, a simple piecewise constant histogram interpolation over a uniformly spaced grid exhibits robustness to discontinuities in unknown locations, with only minor changes in coefficients in response to stretching or compression.[14] Applying this modified criterion, the following guarantee can be provided.

Lemma 14. *Let Algorithm (1) be applied to the model of Section (2.2), with distributions $g()$ and $q()$ each chosen to be at least twice differentiable over bounded support, using a tensor product histogram interpolation scheme with K evenly spaced grid points per dimension.*

Assume the steady state is computed by a method such that the value of the steady state functions on this grid converges at rate $O(K^{-1})$ uniformly over all grid points, and either the steady state parameterized expectation function $\tilde{\ell}^*(w_i)$ is monotone over the grid or that the convergence rate is $o(K^{-1})$.

Assume also that the full system satisfies Condition (2) in Childers [2018]. Let Λ^α be the class of Hölder continuous functions with exponent $\alpha \geq \frac{1}{2}$ Then the resulting approximate solutions satisfy, for $K > \bar{K}$ for some $\bar{K}$

$$\begin{aligned} \sup_{\{f \in \Lambda^\alpha:\ \|f\|_{L^2}=1\}} \left\| (\tilde{h}_x^K - h_x)[f(.)] \right\|_{L^2} &\leq O(K^{-\frac{1}{4}}) \\ \sup_{\{f \in \Lambda^\alpha:\ \|f\|_{L^2}=1\}} \left\| (\tilde{g}_x^K - g_x)[f(.)] \right\|_{L^2} &\leq O(K^{-\frac{1}{4}}) \end{aligned}$$

Further, impulse response functions satisfy

$$\sup_{\{f \in \Lambda^\alpha:\ \|f\|_{L^2}=1\}} \left\| ((\tilde{h}_x^K)^m - (h_x)^m)[f(.)] \right\|_{L^2} \leq O(K^{-\frac{1}{4}})$$

$$\sup_{\{f \in \Lambda^\alpha:\ \|f\|_{L^2}=1\}} \left\| (\tilde{g}_x^K (\tilde{h}_x^K)^m - g_x (h_x)^m)[f(.)] \right\|_{L^2} \leq O(K^{-\frac{1}{4}})$$

Proof. See Appendix. □

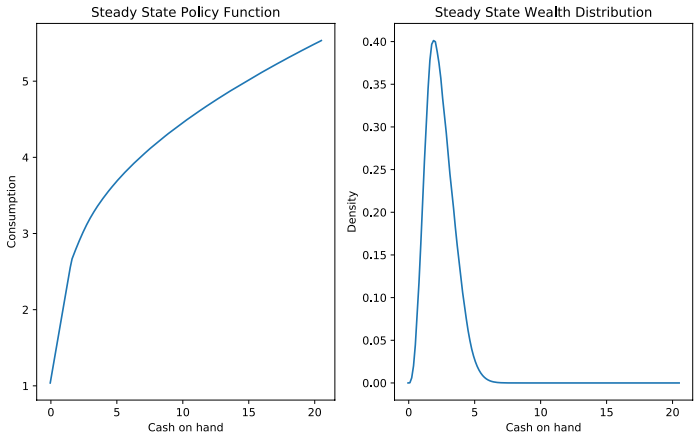
This rate results from the histogram approximation error in uniform norm for piecewise smooth functions with known jumps, which allows application of the bounds of Theorem (10), along with an additional loss due to possible approximation error in the location of the discontinuities, which can be bounded

[14] Strictly, this basis exhibits a form of "local diffeomorphism invariance," in the sense of Bruna and Mallat [2013], Grohs et al. [2018]. See Appendix Section (A.3).

due to local diffeomorphism invariance for appropriately smooth inputs, provided that the steady state approximation satisfies the required conditions and so ensures that the location of the discontinuity is approximated with sufficient accuracy. Because of the need to perform this transformation, the resulting rate is slower than standard function approximation rates for histograms, and applies only over input functions which are sufficiently smooth. Nevertheless, the method is consistent, and numerical results suggest that when using a fine enough grid to locate the discontinuity with sufficient precision, the resulting solutions appear sensible.

In what follows, I display numerical results obtained from applying the method, including steady state values and impulse response functions over 10 periods to 1 unit shocks to ϵ^z and ϵ^σ. The model parameters chosen for the evaluation are chosen only to be illustrative. In all figures and calculations, I set $\underline{a} = -0.55$, $\gamma = 2$, $q(z) = \frac{\bar{z}-\underline{Z}}{2\mathcal{I}} \exp(\frac{-1}{1-(-1.0+\frac{2}{\bar{z}-\underline{Z}}(z-\underline{Z}))^2})$ on $z \in [\underline{z}, \bar{z}]$ and 0 elsewhere, where $\mathcal{I} = 0.443993816237631$ is a constant ensuring that q is a valid density, $\bar{z} = 2$, and $\underline{z} = 0$. I set $g(s)$ as a truncated log normal, with mean parameter 0, variance parameter 1, and lower and upper bounds 0.5 and 3.5 respectively. For the aggregate shock to income, I set persistence $\rho_z = 0.9$. The shock to σ_t has persistence $\rho_\sigma = 0.5$, and enters into the distribution $g(s)$ by multiplying the variance factor by $\exp(\sigma_t)$. Graphs are calculated using $K = 150$ evenly spaced grid points over interval $[-0.5, 20.5]$. Figure (1) displays steady state cash on hand distribution $m^*(w)$ and consumption policy function $c(w, \ell^*(w), R^*)$. Figure (2) displays impulse responses to a 1 unit shock to ϵ^z, the aggregate income. This shock raises consumption for consumers at all levels of the wealth distribution, though the effect is largest for those with wealth just above the level of the borrowing constraint. In equilibrium, the temporary increase in income results in greater demand for savings and so a decline in the market clearing interest rate. The effect on the cross-sectional distribution of wealth is multimodal, increasing the prevalence of high and low wealth while reducing the prevalence of wealth at intermediate levels. Figure (3) displays the impulse response to an anticipated shock to the variance parameter of the truncated log normal distribution. Note that such a shock is not a pure mean preserving spread, due to the parameterization, and raises both mean and variance of the income distribution. On impact, this shock raises the consumption of the borrowing constrained, who anticipate higher income (as the truncation point is unchanged), and slightly lowers that of the unconstrained, with net result an increase in demand for savings which lowers the interest rate. In one period, when the change in the income distribution is realized, consumption falls at all wealth levels, though more for the unconstrained, and demand for savings falls, raising the interest rate above the steady state level. The distributional result of this set of changes is to shift the wealth distribution left, reducing the mass of high wealth and raising the mass of low wealth consumers.

Figure 1: Huggett Model: Steady State



5 Conclusion

Bringing together models of dynamic heterogeneous individual level microeconomic behavior and aggregate macroeconomic variability is a challenge that has admitted substantial interest and computational effort. However, unlike the case of dynamic macroeconomic models with limited heterogeneity, for which there exist modular, general purpose modeling frameworks with rigorous computational guarantees [Fernández-villaverde et al., 2016], methods for heterogeneous agent models, which must handle endogenously varying function valued variables, have tended to be supported mainly by intuition and numerical experiments, or have guarantees provided only in highly restricted model classes. In this work, I have taken steps to provide a modular, extensible framework which nests versions of a broad cross section of dynamic heterogeneous agent models and which possess computational guarantees which should permit extension and experimentation to incorporate novel mechanisms and empirical contexts in a common computational environment. This framework specializes the existing approach of combining perturbation and projection methods to solve heterogeneous agent models to a class of models built from a small set of linear and pointwise nonlinear building blocks and shows how to combine these in a way which can generate standard models and many extensions thereof, and how to apply automated methods to reliably produce a linearized solution.

While the broad outlines of this method are similar to existing approaches,

Figure 2: Impulse response to ϵ^z shock

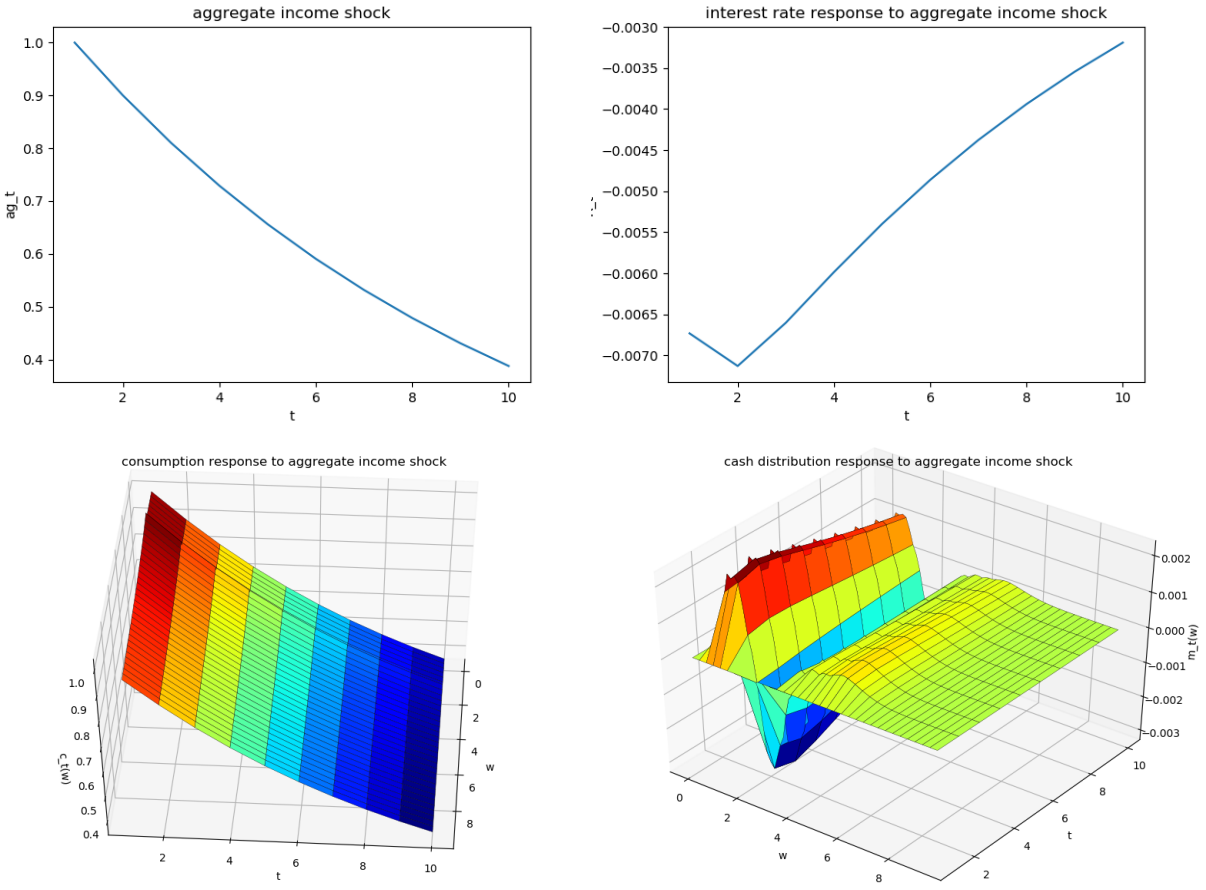
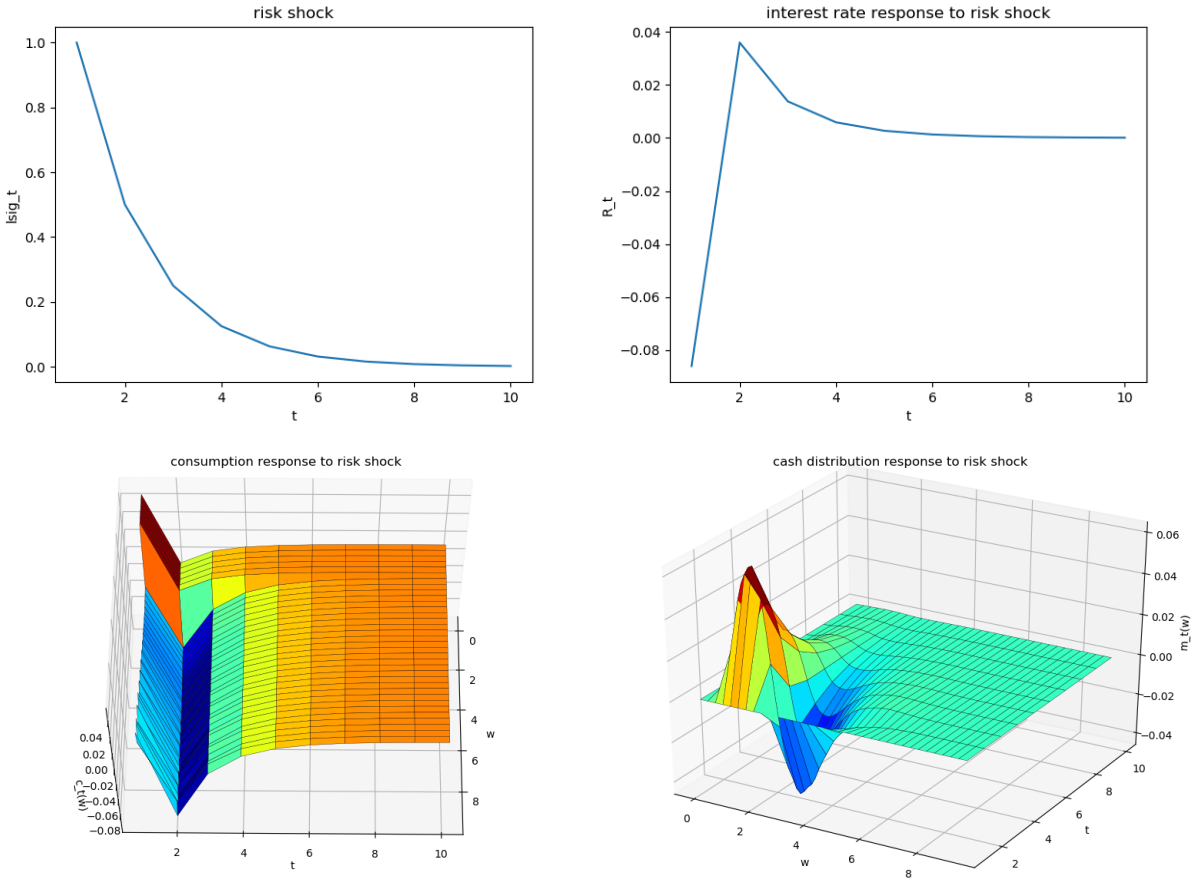


Figure 3: Impulse response to ϵ^{σ} shock



the details of the analysis provided here suggest details can matter substantially for the kinds of guarantees which can be provided. Models with derivatives which are built out of kernel integral operators that approximate the kernel functions by accurate interpolation methods can produce matrix representations with very strong approximation guarantees, while models outside of this class present substantial challenges, remaining outside of the reach of available analyses. In this work, it was demonstrated how models in this tractable class can be built and extended in great generality, and how existing models can be transformed to fit in the class through proper construction of equilibrium conditions and choice of variables. By showing how the form and approximation guarantees can be preserved under composition via computational graphs, it is possible to build models of a high degree of complexity and continue to ensure their membership in a tractable class and so preserve the computational guarantees. This is demonstrated in the general class of heterogeneous agent models of Arellano and Bonhomme [2016] and in the special case of a version of the consumption-savings model of Huggett [1993], but the tools are applicable in a much broader variety of cases, and it is hoped that this generality will promote new uses of heterogeneous agent models to explore novel aspects of consumer, firm, geographic, and other forms of microeconomic heterogeneity.

While the method provided greatly enhances the generality of the classes of models which can be reliably evaluated, there are noteworthy limitations that suggest cases in which alternative methods might be preferred. The ability to ensure strong guarantees for models built out of kernel operators need not imply that models which do not take this structure cannot enjoy strong or superior performance using alternative methods. Heterogeneous agent models in continuous time defined in terms of differential equations do not fit in this class, but a perturbation approach to these models, as in Ahn et al. [2017], may continue to enjoy similar guarantees based on the well developed convergence theory for numerical approximation of differential equations. In discrete time, perturbation methods which do not explicitly transform the model representation to ensure tractable structure, such as those of Reiter [2009] and Winberry [2016] may nevertheless exhibit strong and generalizable performance. Given the broad applicability of these methods in practice, it is hoped that the analysis provided here might also provide a foundation upon which to validate or enhance the performance of these less structured methods.

References

Robert Alexander Adams and John J.F. Fournier. *Sobolev Spaces*. Academic Press, Amsterdam, 2nd ed. edition, 2003.

Ben Adcock, Anders Hansen, Bogdan Roman, and Gerd Teschke. Generalized sampling: Stable reconstructions, inverse problems and compressed sensing over the continuum. *Advances in Imaging and Electron Physics*, 182(viii): 187–279, 2014. ISSN 10765670. doi: 10.1016/B978-0-12-800146-2.00004-7.

SeHyoun Ahn, Greg Kaplan, Benjamin Moll, Thomas Winberry, and Christian Wolf. When Inequality Matters for Macro and Macro Matters for Inequality. Working Paper 23494, National Bureau of Economic Research, jun 2017. URL http://www.nber.org/papers/w23494.

Treb Allen, Costas Arkolakis, and Xiangliang Li. On the existence and uniqueness of trade equilibria. 2015. ISSN 00989886. doi: 10.1214/07-EJS004. URL http://www.econ.yale.edu/ ka265/research/GravityTrade/AA - note - existence and uniqueness gravity trade.pdf.

Manuel Arellano and Stéphane Bonhomme. Nonlinear Panel Data Methods for Dynamic Heterogeneous Agent Models. *Annual Review of Economics*, 2016.

Manuel Arellano, Richard Blundell, and Stéphane Bonhomme. Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework. *Econometrica*, 85(3):693–734, 2017. ISSN 0012-9682. doi: 10.3982/ECTA13795.

Costas Arkolakis, Arnaud Costinot, and Andrés Rodríguez-Clare. New trade models, same old gains? *American Economic Review*, 102(1):94–130, 2012.

Adrien Auclert, Bence Bardóczy, Matthew Rognlie, and Ludwig Straub. Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models. jul 2019. URL http://www.nber.org/papers/w26123.

Francis Bach. On the Equivalence between Quadrature Rules and Random Features. *Journal of Machine Learning Research*, 18:1–38, 2017. doi: HAL-01118276-v2.

Rüdiger Bachmann, Ricardo J. Caballero, and Eduardo M. R. a. Engel. Aggregate implications of lumpy investment: New evidence and a DSGE model. *American Economic Journal: Macroeconomics*, 5(4):29–67, 2013. ISSN 19457707. doi: 10.1257/mac.5.4.29.

Volker Barthelmann, Erich Novak, and Klaus Ritter. High dimensional polynomial interpolation on sparse grids. *Advances in Computational Mathematics*, 12:273–288, 2000. ISSN 1019-7168. doi: 10.1023/A:1018977404843.

A. Berlinet and Christine Thomas-Agnan. *Reproducing Kernel Hilbert Spaces in Probability and Statistics*. Kluwer Academic, Boston, 2004.

G. Beylkin, Ronald R. Coifman, and V. Rokhlin. Fast Wavelet Transforms and Numerical Algorithms I, 1991. ISSN 00103640. URL http://amath.colorado.edu/pub/wavelets/papers/BE-CO-RO-1991.pdf.

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

Timo Boppart, Per Krusell, and Kurt Mitman. Exploiting MIT Shocks in Heterogeneous-Agent Economies: The Impulse Response as a Numerical Derivative. Working Paper 24138, National Bureau of Economic Research, dec 2017. URL http://www.nber.org/papers/w24138.

John P Boyd. *Chebyshev and Fourier Spectral Methods*. Dover, New York, 2 edition, 2000. ISBN 9780486411835.

Johannes Brumm and Simon Scheidegger. Using Adaptive Sparse Grids to Solve High-Dimensional Dynamic Models. *Econometrica*, 85(5):1575–1612, 2017.

Joan Bruna and Stephane Mallat. Invariant scattering convolution networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8): 1872–1886, 2013. ISSN 01628828. doi: 10.1109/TPAMI.2012.230.

Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numerica*, 13(June 2004):147, 2004. ISSN 0962-4929. doi: 10.1017/S0962492904000182. URL http://www.journals.cambridge.org/abstract_S0962492904000182.

Michele Campiti, Giusy Mazzone, and Cristian Tacelli. On the interpolation of discontinuous functions. *Journal of Approximation Theory*, 164(5):731–753, 2012. ISSN 00219045. doi: 10.1016/j.jat.2012.02.003.

Christopher D Carroll. Theoretical foundations of buffer stock saving. Technical report, National Bureau of Economic Research, 2004.

Christopher D Carroll and Miles S . Kimball. On the Concavity of the Consumption Function. *Econometrica*, 64(4):981–992, 1996.

Françoise Chatelin. Numerical Approximation Methods for Integral and Differential Operators. In *Spectral Approximation of Linear Operators*, chapter 4. Society for Industrial and Applied Mathematics, 2011. doi: 10.1137/1.9781611970678.

David Childers. *Computational Methods for Economic Models with Function-valued States*. Phd thesis, Yale University, 2016.

David Childers. Solution of Rational Expectations Models with Function Valued States. 2018.

David Childers and Keshav Dogra. Perturbation Methods for Incomplete Markets Economies. 2018.

Lawrence J. Christiano and Jonas D.M. Fisher. Algorithms for solving dynamic models with occasionally binding constraints. *Journal of Economic Dynamics and Control*, 24(8):1179–1232, 2000. ISSN 01651889. doi: 10.1016/S0165-1889(99)00016-0.

Hess Chung. Perturbative Solutions for Heterogeneous Agent DSGE Models. 2007.

James Costain and Anton Nakov. Distributional dynamics under smoothly state-dependent pricing. *Journal of Monetary Economics*, 58(6-8):646–665, 2011. ISSN 03043932. doi: 10.1016/j.jmoneco.2011.11.006. URL http://dx.doi.org/10.1016/j.jmoneco.2011.11.006.

David L. Donoho. Sparse components of images and optimal atomic decompositions. *Constructive Approximation*, 17(3):353–382, 2001.

David L. Donoho and Emmanuel J. Candès. Continuous curvelet transform: II. Discretization and frames. *Applied and Computational Harmonic Analysis*, 19(2):198–222, 2005.

Tobin A Driscoll, Nicholas Hale, and Lloyd N. Trefethen. *Chebfun Guide*. 2014.

Jesús Fernández-villaverde, Juan F. Rubio-Ramírez, and Frank Schorfheide. Solution and Estimation Methods for DSGE Models. jan 2016. URL http://www.nber.org/papers/w21862.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, 2016.

R.A. Gopinath and C.S. Burrus. On the Moments of the Scaling Function psi_0. *Proc. 1992 IEEE International Symposium on Circuits and Systems*, (2):963–966, 1992.

Philipp Grohs, Thomas Wiatowski, and Helmut Bölcskei. Deep Convolutional Neural Networks on Cartoon Functions. 2018. doi: 10.1109/ISIT.2016.7541482. URL http://arxiv.org/abs/1605.00031 http://dx.doi.org/10.1109/ISIT.2016.7541482.

William H. Guss. Deep Function Machines: Generalized Neural Networks for Topological Layer Expression. 2017. URL http://arxiv.org/abs/1612.04799.

William H. Guss and Ruslan Salakhutdinov. On Universal Approximation by Neural Networks with Uniform Guarantees on Approximation of Infinite Dimensional Maps. pages 1–16, 2019. URL http://arxiv.org/abs/1910.01545.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016. ISSN 1664-1078. doi: 10.1109/CVPR.2016.90. URL http://ieeexplore.ieee.org/document/7780459/.

Jonathan Heathcote, Kjetil Storesletten, and Giovanni L. Violante. Quantitative macroeconomics with heterogeneous households. *Annu. Rev. Econ.*, 1(1):319–354, 2009.

Hugo A. Hopenhayn. Entry, Exit, and Firm Dynamics in Long Run Equilibrium. *Econometrica*, 60(5):1127–1150, 1992.

Kurt Hornik, Maxwell B. Stinchcombe, and Halbert White. Multilayer Feedforward Networks are Universal Approximators. *Neural Networks*, 2:359–366, 1989.

Mark Huggett. The risk-free rate in heterogeneous-agent incomplete-insurance economies. *Journal of Economic Dynamics and Control*, 17(5-6):953–969, 1993.

Iain M. Johnstone. *Gaussian estimation: Sequence and wavelet models*. Unpublished manuscript, 2015.

Iain M. Johnstone and Bernard W Silverman. Boundary Coiflets for Wavelet Shrinkage in Function Estimation. *Journal of Applied Probability*, 41(A (Stochastic Methods and Their Applications)):81–98, 2004.

Kenneth L. Judd, Lilia Maliar, Serguei Maliar, and Rafael Valero. Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. *Journal of Economic Dynamics and Control*, 44:92–123, 2014. ISSN 01651889. doi: 10.1016/j.jedc.2014.03.003. URL `http://dx.doi.org/10.1016/j.jedc.2014.03.003`.

Greg Kaplan, Benjamin Moll, and Giovanni L. Violante. Monetary Policy According to HANK. jan 2016.

S. Kesavan. *Nonlinear Functional Analysis: A First Course*. Hindustan Book Agency, New Delhi, 2004.

Aubhik Khan and Julia K. Thomas. Idiosyncratic shocks and the role of nonconvexities in plant and aggregate investment dynamics. *Econometrica*, 76 (2):395–436, 2008. ISSN 00129682. doi: 10.1111/j.0012-9682.2008.00837.x.

Andreas Klimke and Barbara Wohlmuth. Algorithm 847: spinterp: Piecewise multilinear hierarchical sparse grid interpolation in MATLAB. *ACM Transactions on Mathematical Software (TOMS)*, 31(4):561–579, 2005.

Paul Krugman. Increasing Returns and Economic Geography. *Journal of Political Economy*, 99(3):483–499, 1991. ISSN 0022-3808. doi: 10.1086/261763. URL `http://www.journals.uchicago.edu/doi/abs/10.1086/261763`.

Per Krusell and Anthony a. Smith, Jr. Income and Wealth Heterogeneity in the Macroeconomy. *Journal of Political Economy*, 106(5):867, 1998. ISSN 00223808. doi: 10.1086/250034.

Lars Ljungqvist and Thomas J. Sargent. *Recursive macroeconomic theory*. MIT press, 3rd edition, 2012.

Dougal Maclaurin, David Duvenaud, and Ryan P Adams. Autograd : Effortless Gradients in Numpy. *ICML 2015 AutoML Workshop*, 2015.

Stephane Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, third edition, 2008.

Mathew W McLean, Giles Hooker, Ana-Maria Staicu, Fabian Scheipl, and David Ruppert. Functional generalized additive models. *Journal of Computational and Graphical Statistics*, 23(1):249–269, 2014.

Richard Nickl. *Statistical Theory*. 2013.

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

Jarrett Revels, Miles Lubin, and Theodore Papamarkou. Forward-Mode Automatic Differentiation in Julia. *CoRR*, abs/1607.0, 2016. URL `http://arxiv.org/abs/1607.07892`.

John Rust. Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033, 1987.

Stephanie Schmitt-Grohé and Martín Uribe. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of Economic Dynamics and Control*, 28(4):755–775, 2004. ISSN 01651889. doi: 10.1016/S0165-1889(03)00043-5.

Won-Ki Seo. Cointegrated Density-Valued Linear Processes. pages 1–35, 2017. URL `http://arxiv.org/abs/1710.07792`.

Alex Townsend, Marcus Webb, and Sheehan Olver. Fast polynomial transforms based on Toeplitz and Hankel matrices. 2016. URL `http://arxiv.org/abs/1604.07486`.

Lloyd N. Trefethen. *Approximation theory and approximation practice*. SIAM, 2013.

Marcelo Veracierto. Adverse Selection, Risk Sharing and Business Cycles. *Federal Reserve Bank of Chicago Working Paper*, (October), 2014.

Thomas Winberry. A Toolbox for Solving and Estimating Heterogeneous Agent Macro Models. 2016.

A Bases and Function Spaces: Choices and Guarantees

As described, the approximation algorithm introduced here can be implemented over a wide variety of models, using a wide variety of interpolation schemes using a variety of classes of basis functions. Which interpolation method is chosen can be based on a combination of ease of implementation, speed, and accuracy

over the types of functions which compose the model to be solved. Properties required of such a method are that it have a representation as an interpolation scheme implementable by a matrix transform and that it be *sup norm accurate* over the classes of functions in the model.[15] Accuracy of the algorithm for a given number of basis functions (or, equivalently, number of basis functions needed for a given desired accuracy) is determined by the precise rates of convergence of the function approximation in sup norm. Additional desirable features are the preservation of these accuracy properties with respect to tensor products and over marginals of multivariate functions. These properties are satisfied by a large set of popular approximation methods and their tensor products over many reasonable function classes. In this appendix, for convenience, I collect a variety of popular methods and associated function classes and present their convergence properties. These include, among others, Chebyshev polynomials, histograms, splines, wavelets, and sparse grid (e.g. Smolyak) interpolation schemes. I also discuss specialized methods for functions with discontinuities, which require delicate modification to the general framework. The list is by no means exhaustive, but should cover many of the most commonly used methods and models. As some of these classes of basis functions, including Chebyshev polynomials and splines, at least in their default implementations, are not orthonormal with respect to the standard L^2 inner product, I overview approaches for orthonormalization, as well as for restricting to functions which integrate to 0, which is needed for arguments which are probability densities, in subsection (A.1).

A note should be made regarding some (possibly conspicuous) omissions. The requirement for sup norm accuracy generally omits models whose definition contains unbounded functions. Much of the approximation literature, especially for function classes allowing non-uniform regularity such as Sobolev and Besov classes and Reproducing Kernel Hilbert Spaces (see, e.g., Adams and Fournier [2003], Johnstone [2015], and Berlinet and Thomas-Agnan [2004], respectively), works in L^2 norm for reasons of tradition and convenience, but for the methods presented here this is not enough to ensure consistency. In some cases this omission is a technicality and stronger results may be shown (see e.g. Bach [2017] for (some) Sobolev classes), but in other cases the absence of uniform regularity permits unbounded functions which in general do not admit uniform approximation. The requirement that the interpolation scheme have representation as a matrix also rules out a variety of modern approaches which ensure higher accuracy over certain function classes, including thresholding and (penalized) nonlinear optimization approaches popular for wavelet or neural network function representations or adapive grid interpolation methods [Brumm and Scheidegger, 2017]. This is inherent to the use of a linearized method and so cannot be avoided. However, these methods may be used for computation of approximate steady states, and partial adaptation can be achieved in some cases: see the section on discontinuous functions. Finally, the requirement that input

[15] This can be relaxed very slightly, at the cost of somewhat weaker conclusions: see Subsection (A.3)

and output function spaces be isomorphic (needed for a numerical solution to exist at all) and that function classes and interpolation schemes be preserved under marginalization (needed for automatic construction of derivatives) requires a tensor product approach to multivariate function representation in any case where the model calls for integrating over only a subset of arguments of a function. This rules out default implementations of a variety of asymmetric multivariate approaches such as sparse grid interpolation [Judd et al., 2014] or 2D wavelet thresholding (Mallat [2008], Beylkin et al. [1991]), though symmetric tensor product versions of these methods can be applied to subsets of arguments which are not separated by marginalization.

A.1 Orthonormalization and regularization

One difficulty with approximating a linear operator using an interpolation scheme is that for many interpolation methods with desirable accuracy properties, the basis set Φ_K is not necessarily an orthonormal basis with respect to the $L^2[\mathcal{S}]$ inner product $\langle f(s), g(s)\rangle = \int_{\mathcal{S}} f(s)g(s)ds$. For example, while interpolation at the Chebyshev nodes is near-optimal for certain classes of smooth functions, the Chebyshev polynomials are orthogonal only with respect to a weighted inner product $\langle f(s), g(s)\rangle_{w(s)} = \int_{\mathcal{S}} f(s)g(s)\frac{1}{\sqrt{1-s^2}}ds$. Similarly, while B-splines are convenient for representing a piecewise polynomial interpolation, they are also not orthogonal. To build a matrix representation of an integral kernel operator by approximating it using coefficients with respect to an orthonormal basis, one simply needs to transform the coefficients of the kernel with respect to a non-orthogonal set of functions into an equivalent representation using an orthonormal basis. While computationally efficient methods are available to perform this transformation for, e.g., Chebyshev polynomial bases (see Townsend et al. [2016]), which are recommended in practice in those cases, for convenience in the case of more general interpolation schemes, I provide a simple procedure, based on Cholesky decomposition, for constructing a map to coefficients of a basis for the span of Φ_K orthogonal with respect to a given inner product $\langle, \rangle$. This procedure may be used whenever it is possible to calculate inner products in closed form, but there is no known representation of the functions Φ_K in terms of coordinates with respect to some known basis set (in which case one could merely apply the QR algorithm to these coordinates).

In essence, what this procedure does is define a matrix representation U of the existing functions Φ_K and then performs Gram-Schmidt orthogonalization of this matrix, by the QR algorithm. The overhead of this method is moderate: Cholesky factorization is of roughly cubic complexity in K. Building the Gram matrix requires the ability to compute inner products, which are often available in closed form for many useful function classes, but finding this form will depend on the class. Most importantly, however, once K and Φ_K are fixed, this operation may be performed independently of the parameters of the model, so in an estimation setting or other application involving recomputing the model, this procedure only needs to be performed once.

Algorithm 2 Orthonormalization

Input: K linearly independent functions $\Phi_K = \{\phi_k\}_{k=1}^K$, inner product $\langle , \rangle$
Output: Functions $\Psi_K = \{\psi_k\}_{k=1}^K$ orthonormal w.r.t. $\langle , \rangle$ and linear map J_K such that $\hat{f}_K(s) = \sum_{k=1}^K \hat{f}_k \phi_k(s) = \sum_{k=1}^K (J_K \hat{f})_k \psi_k(s)$, i.e., J_K maps Φ_K coefficients to Ψ_K coefficients

1. Build Gram matrix G: $[G]_{i,j} = \langle \phi_i, \phi_j \rangle$
2. Cholesky decompose G: $U^*U = chol(G)$
3. Construct new basis functions: $X = U^{-1}$, $\psi_k(s) = \sum_{j=1}^K X_{jk} \phi_j(s)$
4. Construct transform to new basis: $J_K = U$

Lemma 15. *Given K linearly independent functions $\Phi_K = \{\phi_k\}_{k=1}^K$ and inner product $\langle , \rangle$ Algorithm (2) produces an orthonormal basis $\Psi_K = \{\psi_k\}_{k=1}^K$ and a linear map J_K such that $\hat{f}_K(s) = \sum_{k=1}^K \hat{f}_k \phi_k(s) = \sum_{k=1}^K (J_K \hat{f})_k \psi_k(s)$ for any $\hat{f}$.*

Proof. To see that Ψ_K are orthonormal with respect to inner product $\langle , \rangle$, note that

$$\begin{aligned} \langle \psi_k, \psi_l \rangle &= \left\langle \sum_{i=1}^K X_{ik} \phi_i(s), \sum_{j=1}^K X_{jl} \phi_j(s) \right\rangle \\ &= \sum_{i=1}^K \sum_{j=1}^K X_{ik} X_{jl} \langle \phi_i, \phi_j \rangle = X_k^* G X_l \\ &= (UX_k)^* (UX_l) \end{aligned}$$

As a result, the Gram matrix of Ψ_K is equal to $X^*U^*UX = U^{-1*}U^*UU^{-1} = I$, and so the basis Ψ_K is orthonormal, as claimed.

To see that J_K gives the desired map between coefficients, note that $\sum_{k=1}^K (J_K X_i)_k \psi_k(s) = \psi_i(s) = \sum_{j=1}^K X_{ji} \phi_j(s)$ for all i. □

In a model which contains state variables which are densities, for example variables f_X and Lf_X in the model of section (2.1), interpolation may require an additional step. In addition to square integrability, required for all function valued variables, densities are restricted to be non-negative and integrate to 1. There are at least two ways to handle this restriction. One approach analogous to log linearization, is to use the unnormalized log density $l_X(X_i)$ as the state variable and replace $f_X(X_i)$ by $\frac{\exp(l_X(X_i))}{\int \exp(l_X(X_i)) dX_i}$ which maintains nonnegativity and normalization by construction; see Seo [2017] and the discussion thereof in Childers [2018]. Alternately, one may continue to use $f_X(X_i)$ as the variable, but restrict to to perturbations in the space $L_0^2(\mathcal{X})$ of functions which integrate

Algorithm 3 Normalization of densities

Input: $\{\mathcal{T}_K, M_K, \Phi_K\}$ be an interpolation scheme over functions on $\mathcal{S}$, with Φ_K orthonormal (without loss of generality).
Output: Interpolation scheme $\{\mathcal{T}_K, S_K M_K, \Psi_K\}$ with orthonormal functions $\Psi_K = \{\psi_k\}_{k=1}^{K-1}$ orthogonal to constant function. $K-1 \times K$ linear map S_K such that such that given $\hat{f}_K(s) = \sum_{k=1}^{K} \hat{f}_k \phi_k(s)$, $\sum_{k=1}^{K-1} (S_K \hat{f})_k \psi_k(s)$ is the Ψ_K representation of the projection of $\hat{f}_K$ onto the orthogonal complement of the constant function.

1. If the constant function $f(s) = 1$ is in Span Φ_K, let v_1 be the coefficient representation of $f(s)$.
   (a) Else, use $v_1 = M_K \overrightarrow{1}$, where $\overrightarrow{1}$ is a $K \times 1$ vector of all 1s.
2. Define P as the $K \times K$ matrix with first column equal to v_1, and all other entries $[P]_{ij} = 1\{i = j\}$, i.e., the $K-1 \times K-1$ identity matrix.
3. $Q_K R_K = qr(P)$ Orthonormalize P by applying QR algorithm.
4. Let $\psi_{k-1} = \sum_{i=1}^{K} [Q_K]_{ik} \phi_i$ for $k = 2 \dots K$
5. Let $S_K = (Q_K[1:K; 2:K])^\top$

to 0 in order to maintain normalization.[16] This restriction can be implemented numerically by interpolating onto a set of basis functions which always integrate to 0. This is achievable by a simple modification of an interpolation approach which maps to functions in L^2, by orthogonalizing the basis with respect to the constant function. Algorithm (3) provides a generic construction.

By applying the above method, one can construct an interpolation matrix $S_K M_K$ which always interpolates to functions which integrate to (approximately) 0. If the original functions are not orthonormal, one can orthonormalize by Algorithm (2) and use matrix $S_K J_K M_K$. In cases where the constant function is within the span of the first K functions, this procedure simply orthogonalizes with respect to this function and removes the function corresponding to the constant function. For orthogonal polynomials, the constant function is already a basis function, and so the approach consists of first applying standard polynomial interpolation, then removing the first basis function. In cases like splines or histograms, where the constant function lies in the span but is not a basis function itself, one applies interpolation then maps to modified versions which have been transformed to integrate to 0. In cases where the constant function is not in the exact span, one induces additional approximation error, but this error is asymptotically of no greater order than $\left\| M_K \overrightarrow{1} - 1 \right\|_\infty$, the in-

[16] This approach maintains non-negativity only for small enough perturbations, requiring a slight reinterpretation of the approximation guarantee: see Childers [2018]

terpolation error in the constant function.[17] For the algorithms provided, it can be shown that the error induced by this approximation will be of lower order and so will not contribute to any of the asymptotic results.

A.2 Function Classes

Chebyshev Polynomials

Chebyshev polynomials provide a good default basis choice for models defined over a bounded interval. They have nearly optimal accuracy properties for representing many common classes of smooth or very smooth functions, are computationally attractive due to fast algorithms for interpolation, integration, and other tasks, and are convenient due to the widespread availability of software implementations. Regarding choice of basis functions for numerical solution of functional equations, Boyd [2000] suggests the following "moral principle":

> (i) When in doubt, use Chebyshev polynomials unless the solution is spatially periodic, in which case an ordinary Fourier series is better.
> (ii) Unless you're sure another set of basis functions is better, use Chebyshev polynomials.
> (iii) Unless you're really, really sure that another set of basis functions is better, use Chebyshev polynomials.

Accordingly, Chebyshev polynomials are recommended as a first choice for models defined in terms of smooth functions. That being said, particular features of a problem in which Chebyshev polynomials are a more apt choice include highly spatially regular functions, low dimensionality, and bounded and rectangular support, and other choices may be necessary or desirable in their absence. Even within the class of polynomial representations, they may yield suboptimal approximations relative to other choices of projection or interpolation, and for finitely differentiable functions slightly better rates may be achieved by wavelets or sometimes splines. However, economic models are often highly regular or even analytic, in which case polynomials, unlike splines or wavelets, achieve exponential convergence rates, and even within the class of finitely differentiable functions, the accuracy cost of Chebyshev polynomials is suboptimal by a logarithmic factor at most (see, e.g. Trefethen [2013, Ch. 7] for a detailed discussion), for which one gains substantially faster algorithms with reliable existing implementations.

Formally, a Chebyshev interpolation scheme may be described as follows

Definition 16. A Chebyshev interpolation scheme of order K is a tuple $\{\mathcal{T}_K, M_K, \Phi_K\}$ consisting of a sequence of point sets $\mathcal{T}_K = \{s_k\}_{k=1}^K \in [-1, 1]$ where s_k are the Chebyshev points of the second kind[18] $s_k = -\cos((k-1)\pi/(K-1))$, linear

[17] Formally, let P_K be the orthogonal projection onto Φ_K and $P_1^\perp$ the projection onto the orthogonal complement of 1 (i..e. $P_1^\perp f = f(s) - \int f(s)ds$). Then a small amount of algebra shows $\left\|S_K P_K + (I - P_K) - P_1^\perp\right\|_{op} \leq O(\left\|M_K \overrightarrow{1} - 1\right\|_{L^2}) \leq O(\left\|M_K \overrightarrow{1} - 1\right\|_\infty)$.

[18] The Chebyshev points of the second kind are used for interpolation of Chebyshev polynomials of the first kind. This (slightly confusing) terminology is entirely standard in this

interpolation maps $M_K \in \mathbb{R}^{K\times K}$ given by the discrete cosine transform of type I, and basis functions $\Phi_K = \{\phi_k(s)\}_{k=1}^K$ given by the Chebyshev polynomials (of the first kind), defined as $\phi_k(s) = \cos((K-1)\cos^{-1}(s))$.

Note. There are many other equivalent definitions of Chebyshev polynomials which may be more efficient to compute. Similarly, although the interpolation map can be represented as a matrix, it is often more convenient to apply the Fast Fourier Transform to compute it. However, for the purpose of Algorithm (1), the explicit matrix construction is needed. In cases where one does not desire to include the boundary in the set of interpolating points, one can use instead the Chebyshev points of the first kind, with essentially similar results. Note that the Chebyshev polynomials do not form an orthonormal basis of $L^2[-1,1]$, and are instead orthogonal with respect to the weighting function $w(s) = \frac{1}{\sqrt{1-s^2}}$. To construct an orthonormal basis with respect to an unweighted inner product, the polynomials need to be transformed from the Chebyshev basis to the Legendre basis via a linear transform J_K. While this could be done using the Gram-Schmidt procedure in algorithm (2), more efficient algorithms already exist for this special case: see Townsend et al. [2016] and its implementation in the software Chebfun [Driscoll et al., 2014]. Regarding the exact quadrature scheme associated with this interpolation scheme, this is precisely Clenshaw-Curtis quadrature, for which accuracy guarantees and fast algorithms for weight construction are already well established: see, e.g. Trefethen [2013, Ch. 19].

The sup norm accuracy of univariate Chebyshev interpolation is extremely well-studied. Typical result is given by Trefethen [2013, Theorem 7.2], quoted below (with changes to notation) for convenience.

Theorem. *Trefethen [2013, Thm. 7.2] Let f and its derivatives through $f^{(v-1)}$ be absolutely continuous on $[-1,1]$ and suppose the ν^{th} derivative $f^{(\nu)}$ is of bounded variation V. Then for any $K \geq v-1$, its Chebyshev interpolant p_K satisfies*

$$\|f - p_K\|_\infty \leq \frac{4V}{\pi v(K+1-v)^v}$$

The hypothesis of bounded variation can be weakened slightly to the assumption that $f^{(v-1)}$ is Lipschitz continuous, in which case the interpolation has accuracy of order $O(K^{-v}\log K)$ (as follows from the remarks following Theorem 7.2 and the Lebesgue constant bound in Trefethen [2013, Thm. 15.3]). This latter bound is classical, while the former may provide tighter results in the common case where the lack of differentiability arises from a small set of kink points.

In cases where infinite numbers of derivatives exist, these polynomial rates can be improved to exponential. Theorem 8.2 in Trefethen presents the bound $\frac{4M\rho^{-(K+1)}}{\rho-1}$, where $|f(s)| \leq M$ and $\rho > 1$ is the radius of the Bernstein ellipse over which the function can be analytically continued. In practice, this means that if the function is bounded and analytic over $[-1,1]$, the interpolation accuracy is

literature.

exponential, with rate depending on the size of the region in the complex plane over which the function is analytic.

Multivariate Polynomial Interpolation

In the case of multivariate functions, there are a number of ways to extend Chebyshev interpolation, with tradeoffs in computational and approximation properties depending on the dimension, the class of functions one desires to approximate, and potentially the algorithm used. The simplest and most general multivariate interpolation method based on the Chebyshev polynomials is to take the tensor product of univariate schemes. Tensor product Chebyshev interpolation achieves achieves sup norm accuracy over classes of finitely differentiable functions comparable to the sum of the univariate rates. If all dimensions are comparable, this results in convergence rates in terms of total number of function evaluations which slow down exponentially with dimension, the well-known curse of dimensionality. Mason (1980, Theorem 4.1) provides precise convergence rates, summarized below.

Lemma. *Let $\{K_z\}_{z=1}^d$ be cardinalities of the interpolation set over $[-1,1]^d$ for Tensor product Chebyshev interpolation scheme $\mathcal{Q}_d := \{\times_{z=1\ldots d}\mathcal{T}_{K_z}, \otimes_{z=1\ldots d}M_{K_z}, \otimes_{z=1\ldots d}\Phi_{K_z}\}$. Let $\mathcal{F}^\alpha = \{f\ [-1,1]^d \to \mathbb{R} :\ \omega_z(t) := \sup_{|s_z - s_z^*|\leq t} |f(s_1,\ldots,s_z,\ldots s_d) - f(s_1,\ldots,s_z^*,\ldots s_d)| \leq c_z t^{-\alpha_z}$ for some $\alpha_z > 0$, $c_z < \infty$ $\}$ be a class of multivariate functions. Then Q_d is sup norm accurate over $\mathcal{F}^\alpha$ at rate $O((\frac{2}{\pi})^d \prod_{z=1}^d \log K_z \cdot \sum_{z=1}^d \omega_z(\frac{1}{K_z+2}))$.*

A straightforward extension of the proof to the case of analytic functions would replace ω_z with an exponentially declining bound.

While unrestricted tensor product bases may be simple to implement, they do not in general result in efficient polynomial approximation for common multivariate function classes. One choice that has recently attracted particular attention in computational economics is interpolation via sparse Smolyak grids, which may be constructed from a carefully chosen nested subset of the tensor product of Chebyshev points. As demonstrated in Barthelmann et al. [2000], Smolyak interpolation leads to substantially milder dependence on the dimension than unrestricted tensor product methods while still achieving fast (albeit not quite rate optimal) convergence for appropriately smooth functions, and for this reason is commonly used in problems requiring approximation of moderately high-dimensional ($d \approx 10 - 20$) functions: see [Brumm and Scheidegger, 2017, Judd et al., 2014]. In particular, Barthelmann et al. [2000] provide guarantees for the class $F_d^k := \{f : [-1,1]^d \to \mathbb{R} : D^j f$ continuous if $j_i \leq k$ for all $i\}$ of functions whose derivatives up to a multi-index of order k in all entries are continuous equipped with norm $\|f\|_{F_d^k} = \max\{\left\|D^j f\right\| : j \in \mathbb{N}_0^d, j_i \leq k\}$.

One difficulty with such schemes for algorithms of this type is that input and output functions must be represented symmetrically, necessitating tensor product representations over at least a subset of dimensions. In order to use a Smolyak approximation method, one would need to use tensor products of Smolyak approximations across different dimensions, reducing the efficiency

gains achievable. In moderate to high dimensions this may still represent a preferable option.

Histograms

Histogram approximation may be the simplest approximation scheme: up to rescaling, it is equivalent to pure discretization, and so provides a theoretical foundation for discretization approaches. While on a theoretical level, it can be treated as simply a special case of wavelet approximation using the scaling functions for Haar wavelets and a one-point quadrature scheme, this method achieves accuracy over a non-periodic domain without special boundary adjustments, and so is simpler to implement in this case.

Definition. A *histogram approximation* is an interpolation scheme $(\mathcal{T}_K, \Phi_K, M_K)$ over domain normalizable to $\mathcal{S} = [0,1) \subset \mathbb{R}$ such that $\mathcal{T}_K = \{\frac{k-\frac{1}{2}}{K}\}_{k=1}^K$, $\Phi_K = \{\phi_k(s)\}_{k=1}^K$, $\phi_k(s) := \sqrt{K}\mathbf{1}\{s \in [\frac{k-1}{K}, \frac{k}{K})\}$ and $M_K = \frac{1}{\sqrt{K}} I_K$, the $K \times K$ identity matrix divided by $\sqrt{K}$.

Remark. Φ_K can easily be seen to be orthonormal with respect to the inner product on $L^2[0,1)$ by the disjoint supports and the $\sqrt{K}$ normalization. The above defines a centered approximation mapping the midpoint of each interval to the associated basis function. One could just as easily use the left or right endpoints (or any other location within the interval) if desired: this will not affect the rate of convergence for approximation of smooth functions, though it may affect constants, or lead to different results in the presence of discontinuities. It is traditional in the wavelet literature to restrict K to an integer power of 2, but this is not needed here; in this interpretation, note that Φ_K correspond to the rescaled scaling functions of the Haar basis at level $L = \log_2 K$: see Nickl [2013].

Lemma 17. *Let $\mathcal{F} = \Lambda^\alpha([0,1)^d)$, a class of Hölder continuous function f on $[0,1)^d$ with Hölder exponent α. Let $\mathcal{Q}_d = \{\times_{z\in\{1\ldots d\}}\mathcal{T}_z, \otimes_{z\in\{1\ldots d\}}\Phi_z, \otimes_{z\in\{1\ldots d\}}M_z\}$ be the d^{th} order tensor product of identical histogram approximations of order K_z, with total cardinality $K = K_z^d$. Then Q_d is sup norm accurate over class $\mathcal{F}$ at rate $\epsilon_K = O(d^{\frac{\min\{\alpha,1\}}{2}} K^{\frac{-\min\{\alpha,1\}}{d}})$.*

Remark. This approximation rate should be compared to the rate for wavelet approximations with higher order regularity properties, which achieve accuracy $O(K^{-\frac{\alpha}{d}})$, which is an improvement if the function has more than 1 derivative, and identical otherwise. This result is based on local approximation of the function over the finite support of the basis functions, as is standard for wavelet representations, but due to the absence of vanishing moments, cannot take advantage of higher order Taylor expansions regardless of the number of derivatives.

Wavelets

Wavelets are a class of basis functions designed to have desirable computational and approximation-theoretic properties by trading off spatial and frequency domain localization. They perform especially well in cases where the functions of interest are of intermediate regularity, with a finite number of derivatives, or have fine spatially localized features. Examples of this sort arise in geostatistical models and (less frequently used in economics) real world audio or image data. Coiflet wavelets were designed by Ingrid Daubechies to satisfy a number of additional properties that make them an ideal basis for numerical approximation of integral equations. Most notably, this includes the use of compactly supported scaling functions with vanishing moments. This permits the use of a one point quadrature scheme which allows the use of a rescaled identity matrix as interpolating transform while maintaining optimal approximation rates for smooth functions. Over a more limited range of regularity scales, these properties also apply to Daubechies wavelets [Gopinath and Burrus, 1992], which also benefit from improved constant factors. The main disadvantage is that these rates apply only over periodic function classes. When performing linear interpolation, it suffices to construct a representation in terms of the scaling functions corresponding to the wavelet class at the finest scale; this permits, as in the histogram case, the use of discretized function values as arguments. However, accuracy guarantees are substantially stronger when the functions are appropriately smooth. The following guarantees are repurposed from Childers [2018].

Definition. *Linear wavelet scaling function interpolation* an interpolation scheme $(\mathcal{T}_K, \Phi_K, M_K)$ over domain normalizable to $\mathcal{S} = [0,1) \subset \mathbb{R}$ such that $\mathcal{T}_K = \{t_k\}_{k=1}^K = \{\frac{k-\frac{1}{2}}{K}\}_{k=1}^K$, $\Phi_K = \{\phi_k(s)\}_{k=1}^K$, and $\phi_k(s) = \sqrt{K}\phi(K(s - t_k - c))$ is a scaling function with s vanishing moments rescaled and recentered to scale $\frac{1}{K}$, where c is a constant corresponding to the center of mass of the particular wavelet[19] and $M_K = \frac{1}{\sqrt{K}} I_K$, the $K \times K$ identity matrix divided by $\sqrt{K}$.

Lemma. *Let $\mathcal{F} = \Lambda^{\alpha}_{per}([0,1)^d)$, a class of periodic Hölder continuous function f on $[0,1)^d$ with Hölder exponent α. Let $\mathcal{Q}_d = \{\times_{z\in\{1\ldots d\}}\mathcal{T}_z, \otimes_{z\in\{1\ldots d\}}\Phi_z, \otimes_{z\in\{1\ldots d\}}M_z\}$ be the d^{th} order tensor product of identical wavelet scaling function approximations of order K_z with s vanishing moments, with total cardinality $K = K_z^d$. Then Q_d is sup norm accurate over class $\mathcal{F}$ at rate $\epsilon_K = O(K^{\frac{-\min\{\alpha,s\}}{d}})$*

Remark. The number of vanishing moments of the scaling function determines the maximum degree of regularity of function at which wavelet approximation achieves optimal accuracy. Coiflets define a family of different scaling functions of increasing regularity, so one can define a Coiflet scaling function with s vanishing moments for any positive integer s. Daubechies wavelets likewise define a family, for which $s = 2$ holds for all orders in the family from $db2$ on (i.e.,

[19] See Beylkin et al. [1991] for this value for Coiflets, and Gopinath and Burrus [1992] for this value for Daubechies wavelets.

for all Daubechies wavelets except the Haar wavelet, which is equivalent to the histogram estimator, with guarantees as above) [Gopinath and Burrus, 1992].

The requirement of periodicity is important for these rates, limiting the class of applications to which they apply. If one is willing to abandon the exact identity transform in favor of a "preconditioning matrix" which takes the form of the identity away from the edge of the domain, the requirement of periodicity can in principle be resolved through the use of boundary Coiflets [Johnstone and Silverman, 2004], which maintain the same accuracy properties with a slightly modified matrix transform M_K. However, the existence of such a basis was demonstrated non-constructively, so implementation remains impractical. More complicated multi-point quadrature schemes may also be possible, at higher computational cost. Mirroring may be used as an extension instead of periodic extension to preserve continuity, albeit rarely does this preserve differentiability. For the case of wavelet methods applied to discontinuous functions (as might occur if periodicity fails), see the section on discontinuous functions below. In these cases, wavelets are still useful, but optimal accuracy might not be achieved except through the use of nonlinear methods.

Splines (Tensor Product and Sparse Grid)

I consider multivariate piecewise polynomial (spline) interpolation using both an unrestricted tensor product basis and a novel basis consisting of a tensor product of interpolations calculated over a sparse grid. Notation and definitions follow Bungartz and Griebel [2004], henceforth BG: see that paper for definitions of all terms. The accuracy of these approximations is calculated for the function class $u \in X_0^{\infty,p+1}([0,1]^d)$, the space of functions in $C^{p+1}([0,1]^d)$ with compact support and with $\|D^\alpha u\|_\infty < \infty$ for any multi-index α with $|\alpha|_\infty \leq p+1$, associated with seminorm $|u|_{p+1,\infty} = \|D^\alpha u\|_\infty$ for $\alpha_j = p+1$ for $j = 1 \ldots d$. While BG allow splines of different order in different dimensions, and the analysis could be extended to that case, for simplicity I consider splines of order p in all directions (ie, if $p = 2$, the functions are piecewise quadratic). The interpolation of u may be represented in terms of the hierarchical Lagrange decomposition specified in that paper, which represents $u \in X_0^{\infty,p+1}([0,1]^d)$ by the decomposition into layers of interpolation over finer and finer grids $u(x) = \sum_\ell u_\ell^{(p)}(x)$, $u_\ell^{(p)}(x) = \sum_{i \in \mathcal{I}_\ell} v_{\ell,i}^{(p)} \phi_{\ell,i}^{(p)}(x)$ where $\phi_{\ell,i}^{(p)}$ is a tensor product of polynomials up to order p (see BG for exact definition) over a bounded support of size $2 \cdot 2^{-\ell}$ in each dimension centered on a subset of the grid points $x_{\ell,i} := i \cdot 2^{-\ell}$, $0 \leq i \leq 2^\ell$, corresponding to indices $i \in \mathcal{I}_\ell := \{\mathbf{i} \in \mathbb{N}^d : \mathbf{1} \leq \mathbf{i} \leq \mathbf{2^l} - \mathbf{1}\ i_j \text{ odd } \forall 1 \leq j \leq d\}$ and $v_{\ell,i}^{(p)}$ are the coefficient values of these basis functions.

Lemma 18. *Let $u \in X_0^{\infty,p+1}([0,1]^d)$ and let $u_n^{p,\infty}$ be the tensor product (hierarchical Lagrange) interpolation of order p of u constructed from values at $(2^n-1)^d$ grid points with rectangular mesh size 2^{-n}. Then $\|u - u_n^{p,\infty}\|_\infty \leq d(\frac{0.5585}{2^{p+1}-1})^d c(p)|u|_{p+1,\infty}(2^{p+1})^{-n}$ where $c(p) = (\frac{2^{p\cdot(p+1)/2}}{(p+1)!})^d$ and so with N func-*

tion evaluations, sup norm error is bounded by $\epsilon_K = d(\frac{0.5585}{2^{p+1}-1})^d c(p)|u|_{p+1,\infty} N^{-\frac{p+1}{d}}$, *or, equivalently,* ϵ *error requires* $O(\epsilon^{-\frac{d}{p+1}})$ *operations.*

Remark. The proof of this result essentially copies the proof of Lemma 3.5 in BG which proves this result for $p = 1$. Despite the restricted function class here, tensor product approximation has complexity identical (up to constants) to tensor product wavelet approximation for Hölder-smooth functions with the same number of derivatives in each direction, even though it assumes stronger restrictions on the cross-derivatives. As a result, for this function class an unrestricted tensor product approximation is clearly suboptimal.

In general, for higher dimensions, for this function class, performance improvements can be achieved by choosing a space other than an unrestricted tensor product space. Intuitively, the cross partial derivatives are smooth up to a high order, so interactions can be approximated well with fewer basis functions. It is therefore economical to approximate using a relatively larger number of functions in each direction, but omit the product of higher order functions, which will be small. For the purpose of operator approximation, tensor products of the bases defining the input functions and the output functions are still useful, as they allow exact representation of the identity over a projected space, but for input spaces which are themselves multidimensional, it is possible and desirable to construct these spaces using sparse approximations. For input spaces of dimension 2 or greater (and so operators with kernels of dimension 4 or greater), substantial savings may be achieved by using a construction which takes the tensor product of two identical sparse grid representations.

Following on BG, who note that an optimal (with respect to sup norm) sparse grid interpolation of $u \in X_0^{\infty,p+1}([0,1]^d)$ has hierarchical basis representation $u_n^{p,1} = \sum_{|\ell|_1 \leq n+d-1} u_\ell^{(p)}$, in contrast with an unrestricted tensor product interpolation which has representation $u_n^{p,\infty} = \sum_{|\ell|_\infty \leq n} u_\ell^{(p)}$, for d even, if I split the multi-index evenly into dimensions $\ell = (\ell^{d_1}, \ell^{d_2})$, $d = d_1 + d_2$, $d_1 = d_2$, an interpolation based on the tensor product of piecewise polynomials of order p over identical sparse grids over $[0,1]^{d_1}$ and $[0,1]^{d_2}$ will have hierarchical basis representation $u_n^{p,\{1,1\}} = \sum_{\max\{|\ell^{d_1}|_1, |\ell^{d_2}|_1\} \leq n+\frac{d}{2}-1} u_\ell^{(p)}$. By adapting the proof of Theorem 3.8 and Lemma 4.7 in BG to this case, it is possible to construct sup norm bounds on the approximation accuracy of this scheme.

Lemma 19. $u \in X_0^{\infty,p+1}([0,1]^d)$ *and let* $u_n^{p,\{1,1\}}$ *be the tensor product over* d_1 *and* d_2 *of sparse grid (hierarchical Lagrange) interpolations of order* p *of* u *constructed as in BG with minimum rectangular mesh size* 2^{-n}. *Then* $\left\| u - u_n^{p,\{1,1\}} \right\|_\infty \leq 0.5585^d c(p)|u|_{p+1,\infty} 2^{-(p+1)n} \cdot 4(\frac{1}{2^{p+1}-1})^{\frac{d}{2}} \cdot A(\frac{d}{2}, n)$ *where* $c(p) = (\frac{2^{p\cdot(p+1)/2}}{(p+1)!})^d$ *and* $A(d,n) = \sum_{k=0}^{d-1} \begin{pmatrix} n+d+i-1 \\ d-1 \end{pmatrix} = \frac{n^{d-1}}{(d-1)!} + O(n^{d-2})$, *and so with* N *function evaluations, sup norm error is bounded by* $\epsilon_N = O(N^{-(p+1)/2} |\log_2 N|^{(p+2)(\frac{d}{2}-1)})$, *or, equivalently,* ϵ *error requires* $O(\epsilon^{\frac{-2}{p+1}} |\log_2 \epsilon|^{\frac{(p+2)(d-2)}{p+1}})$ *operations.*

Remark. As can be seen from the exponents, when $d = 2$, the order bounds achieved here are identical to those in the previous case, as they should be since the schemes are identical in that case. For larger d, substantial savings are achieved: the bound goes from requiring a number of operations for fixed error level depending on N to an exponent in d to a number only taking $\log N$ to an exponent in d, which is characteristic of sparse grids methods. In comparison to the results in BG, which allow sparsification in all dimensions, the tensor product of sparse grids construction achieves slower rates: order $N^{-(p+1)/2}$ instead of order $N^{-(p+1)}$ up to logarithmic terms. This is equivalent to the loss of a tensor product scheme in 2 dimensions relative to a sparse grid method: however, as dimension grows, the difference does not get worse. This is as should be expected, as additional dimensions beyond the second incorporate sparse construction. While the main reason for requiring identical grids in input and output variables in this fashion is to permit representation of input and output functions with identical accuracy, one may conjecture that this construction, which incorporates more grid points along the interaction of input and output variables, but fewer within the class, may also achieve comparable rates for a less restricted function class, with fewer cross-partial derivatives across input and output spaces, analogous to the way full tensor product methods attain optimal rates for Hölder classes, but such an extension will not be pursued here.

Combining these results, tensor product sparse grid spline interpolation up to level n forms an interpolation scheme with $K = \sqrt{N} = \sum_{i=0}^{n-1} 2^i \cdot \begin{pmatrix} d-1+i \\ d-1 \end{pmatrix}$ points, with $\mathcal{T}_K = \{(x_{\ell^{d_1},i}, x_{\ell^{d_2},j}) : \ i \in \mathcal{I}_{\ell^{d_1}}, \ j \in \mathcal{I}_{\ell^{d_2}}, \text{ for } |\ell^{d_1}|_1 \leq n + d_1 - 1, |\ell^{d_2}|_1 \leq n + d_2 - 1\}$ the points, M_K the hierarchical Lagrangian interpolation defined in BG section 4.2, and $\Phi_K = \{\phi^{(p)}_{\ell^{d_1},i}(s)\phi^{(p)}_{\ell^{d_2},j}(t) : \ i \in \mathcal{I}_{\ell^{d_1}}, \ j \in \mathcal{I}_{\ell^{d_2}}, \text{ for } |\ell^{d_1}|_1 \leq n + d_1 - 1, |\ell^{d_2}|_1 \leq n + d_2 - 1\}$ with sup norm accuracy for class $X_0^{\infty,p+1}([0,1]^d)$ at rate $\zeta_K = O(K^{-(p+1)} \left|\log_2 K\right|^{(p+2)(\frac{d}{2}-1)})$.

The above results can be extended beyond the spaces $X_0^{\infty,p+1}([0,1]^d)$ with known values on the boundary (normalized without loss of generality to 0) to the slightly larger class $X^{\infty,p+1}([0,1]^d)$ with unrestricted boundary values, at the cost of introducing interpolation points at the boundaries. For unrestricted tensor products, this results in a change from $(2^n - 1)^d$ grid points to $(2^n + 1)^d$ grid points, changing constants but leaving asymptotic order the same. For sparse grids, on the boundary one may use the grid points corresponding to a sparse grid of the same order but of dimension corresponding to the dimension of the boundary. Following BG (3.77) and Lemma 3.6, this adds

$$s(d) = \sum_{j=0}^{d} \begin{pmatrix} d \\ j \end{pmatrix} 2^{d-j} 2^n \left(\frac{n^{j-1}}{(j-1)!} + O(n^{j-2})\right)$$

additional grid points for a grid of order n in dimension d, so for a tensor product of such grids as in Lemma (19), $(s(\frac{d}{2}))^2$ grid points are added. This again does not affect the rate, but increases the constant exponentially in dimension, so such methods are often impractical in moderate dimensions, as demonstrated in

Klimke and Wohlmuth [2005]. It is therefore suggested that the scheme without boundary points may be used even in the absence of fixed boundary conditions, provided that internal points may be accurately extrapolated to the edges of the state space. For many economic applications, in which the object of interest is an integral equation, this may be a reasonable assumption. An exception is in hard constraints at known locations, for which it is generally crucial to use a piecewise approximation with break at the location of the constraint.

More generally, at the interpolation stage, it may be preferable to use adaptive sparse grids, in which the hierarchical surpluses are computed sequentially along a tree until a desired error tolerance is reached. While this method can produce substantial savings in many practical problems [Brumm and Scheidegger, 2017], existing adaptive schemes would require modification in order to preserve tensor product structure, and to work inside of estimation procedures where functions may need to be approximated across different parameter values, and so a complete analysis of such methods is left as a promising extension.

A.3 Discontinuous Functions

In some cases, notably in the presence of hard borrowing constraints which result in equilibrium conditions which are not continuously differentiable and so are functions with discontinuities in the derivatives, none of the above function classes or interpolation methods ensures uniform convergence. In the case with a finite set of discontinuities with location which is known exactly and regularity over the regions between the discontinuities, there is a very simple solution. Simply apply sup norm accurate approximation methods for regular functions piecewise over each region of regularity. The approximation rate is then the minimum of the approximation rate in each region. In the one-dimensional case, these regions are intervals and so any of the above methods and guarantees may be used without modification. In the multidimensional case this may result in non-rectangular domains: in these cases, approaches based on piecewise polynomials (referred to as Finite Element Methods in the differential equations literature) and certain classes of wavelets [Mallat, 2008] exist which provide similar guarantees.

The case where the location of a discontinuity is not known exactly presents a greater challenge, in that, in general, sup norm error of a magnitude comparable in size to the jump may be difficult or impossible to avoid. In these cases, convergence guarantees are still feasible but cannot be based on sup norm accuracy and so require another approach. In what follows I introduce two related but distinct methods for handling this kind of discontinuity, with advantages and disadvantages of each. The first takes advantage of the fact that the bound on convergence rates in terms of uniform convergence of kernel functions is not quite tight: there exist nontrivial classes of functions with discontinuities and associated approximation schemes which do not converge uniformly but do converge in a sense which is sufficient to guarantee convergence of operator norms

of functional derivatives and so operator norm convergence of a solution. The disadvantage of such an approach is that the types of discontinuities which can be handled by this method are limited, ruling out a variety of models of practical interest, including, notably, the Huggett model introduced here as they have been set up. For this reason, I also propose an alternate approach, which can handle a broader variety of models, at the cost of only ensuring convergence of a solution in a weaker sense, with uniformity not over the full class of square integrable input functions but only over a subclass of smooth input functions. The idea behind this approach is that one can compute an approximation of the derivatives as if using a basis in which the location of the discontinuity is known exactly, so long as one only attempts to evaluate the results using functions for which approximation in a known basis provides an accurate approximation with respect to the one in which the location is known.

Relaxing uniform convergence

Although convenient for analysis, sup norm convergence is a sufficient rather than a necessary condition for convergence of the functional derivatives in operator norm, and can be weakened slightly. To be precise, Young's inequality (Johnstone [2015] Thm. C.26), which provides an upper bound on the operator norm error of an approximated integral operator in terms of the approximation of the corresponding kernel function, does not require uniform convergence of the kernel, but can be weakened to depend on slightly less restrictive set of norms.

Definition. Consider a function $f(x,y)$ $\mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}$. Define the norms $\|f(x,y)\|_{1/\infty} := \text{ess. } \sup_y \int |f(x,y)|\, dx$ and $\|f(x,y)\|_{\infty/1} := \text{ess. } \sup_x \int |f(x,y)|\, dy$ where dx and dy refer to Lebesgue measure over $\mathbb{R}^{d_x}$ and $\mathbb{R}^{d_y}$, respectively.

Using this notation, if the operator is of the form $\int K(x,y)[.]dy$ over bounded subsets of $\mathbb{R}^d$ and the approximation is $\int \hat{K}(x,y)[.]dy$, a refined version of Young's inequality bounds the operator norm error (with respect to inputs in $L^2(dy)$) by

$$\begin{aligned} M_1^{\frac{1}{2}} M_2^{\frac{1}{2}} &:= \left\|\hat{K}(x,y) - K(x,y)\right\|_{1/\infty}^{\frac{1}{2}} \left\|\hat{K}(x,y) - K(x,y)\right\|_{\infty/1}^{\frac{1}{2}} \qquad (32) \\ &\leq C \left\|\hat{K}(x,y) - K(x,y)\right\|_{\infty} \end{aligned}$$

If M_1 or M_2 can be shown to converge directly, sup norm convergence can be dispensed with. This quantity can be controlled for some but not all functions with discontinuities using appropriate classes of interpolation map. For this task, for any $p < \infty$, convergence in L^p is not sufficient (ruling out direct application of much of the numerical analysis literature based on these norms), but the gap between convergence of $M_1^{\frac{1}{2}} M_2^{\frac{1}{2}}$ and of $\left\|\hat{K}(x,y) - K(x,y)\right\|_{\infty}$ can be decisive in some cases, as the following example demonstrates.

Example. Consider kernel function $K(x,y) = 1\{x \geq y\}$ over $[0,1]^2$. Then an evenly spaced histogram approximation using K grid points in each dimension will not converge in sup norm due to the discontinuity along the line segment $x = y$ which ensures that some cells will overlap the discontinuity, resulting in an error of constant order. In this case, however, M_1 and M_2 can be shown to converge at rate $\frac{1}{K}$ as the measure of the region along which pointwise error does not go to 0 is of this order for every x and y.

This simple example can be extended to much broader classes of functions with similar discontinuities, and particularly so for approximation methods which are local and so constrain the region over which pointwise error occurs to a shrinking subspace. For example, the class of cartoon functions [Donoho, 2001] and basis function classes like curvelets [Donoho and Candès, 2005] were specifically designed for representation of functions with discontinuities which are "not too complicated" in a formal sense.

To demonstrate that the procedure defined in Algorithm (1) can accommodate this class, I propose a modified set of conditions and modify the convergence bounds of Lemma (9) accordingly. This will require construction of bounds on the $\|.\|_{1/\infty}$ and $\|.\|_{\infty/1}$ norm approximation errors of the node functions in cases where sup norm accuracy may not be achievable.

As errors are propagated along the nodes of the computational graph defined by the model, this is achieved by replacing the uniform approximation error bounds on the intermediate nodes with bounds over a norm which requires uniformity only over the subset of arguments where this is required for $\|.\|_{\infty/1}$ or $\|.\|_{1/\infty}$ norms over the final nodes. To that end, define the following set of function norms, in terms of the sets of arguments defined for each node in Condition (6). I will call an interpolation scheme $\{\mathcal{T}_K, M_K, \Phi_K\}$ accurate in norm $\|.\|_a$ at rate ϵ_K for function class $\mathcal{F}$ if $\forall f \in \mathcal{F}$, $\left\|\hat{f}_K(s) - f(s)\right\|_a \leq \epsilon_K$.

Definition. $\forall \ell = 1 \dots d_1$, $\forall p^\ell = 1 \dots P^\ell$, $\forall j = 1 \dots 2d_2 + o^\ell$, $\forall m \in M_j^{p^\ell}$ define

$$\|f(.,.,.)\|_{p_j^{\ell m},a} = \underset{s_{[p_j^{\ell m}]/([p_j^{\ell m}]\cap[\ell^o])} \in \mathcal{S}_{[p_j^{\ell m}]/([p_j^{\ell m}]\cap[\ell^o])}}{\text{ess. sup}} \int \left| f(s'_{[p_{ej}^{\ell m}]}, t_{([p_j^{\ell m}]\cap[\ell^o])}, s_{[p_j^{\ell m}]/([p_j^{\ell m}]\cap[\ell^o])}) \right| ds'_{[p_{ej}^{\ell m}]} dt_{([p_j^{\ell m}]\cap[\ell^o])}$$

$$\|f(.,.,.)\|_{p_j^{\ell m},b} = \underset{t_{([p_j^{\ell m}]\cap[\ell^o])} \in \mathcal{S}_{([p_j^{\ell m}]\cap[\ell^o])}}{\text{ess. sup}} \int \left| f(s'_{[p_{ej}^{\ell m}]}, t_{([p_j^{\ell m}]\cap[\ell^o])}, s_{[p_j^{\ell m}]/([p_j^{\ell m}]\cap[\ell^o])}) \right| ds'_{[p_{ej}^{\ell m}]} ds_{[p_j^{\ell m}]/([p_j^{\ell m}]\cap[\ell^o])}$$

$\forall \ell = 1 \dots d_1$, $\forall p^\ell = 1 \dots P^\ell$, $\forall j = 1 \dots 2d_2 + o^\ell$, define also

$$\|f(.,.,.)\|_{p_j^{\ell},a} = \underset{s \underset{m \in M_j^{p^\ell}}{\cup} [p_j^{\ell m}]/([p_j^{\ell m}]\cap[\ell^o]) \in \mathcal{S}}{\text{ess. sup}} \int \left| f(s' \underset{m \in M_j^{p^\ell}}{\cup} [p_{ej}^{\ell m}], t \underset{m \in M_j^{p^\ell}}{\cup} ([p_j^{\ell m}]\cap[\ell^o]), s \underset{m \in M_j^{p^\ell}}{\cup} [p_j^{\ell m}]/([p_j^{\ell m}]\cap[\ell^o])) \right| ds \underset{m \in M_j^{p^\ell}}{\cup} [p_{ej}^{\ell m}] dt \underset{m \in M_j^{p^\ell}}{\cup} ([p_j^{\ell m}]\cap[\ell^o])$$

$$\|f(.,.,.)\|_{p_j^{\ell},b} = \underset{t \in \mathcal{S}}{\text{ess. sup}}$$

$$\int \left| f(s' \underset{m \in M_j^{p^\ell}}{\cup} [p_{ej}^{\ell m}], t \underset{m \in M_j^{p^\ell}}{\cup} ([p_j^{\ell m}] \cap [\ell^o]), s \underset{m \in M_j^{p^\ell}}{\cup} [p_j^{\ell m}]/([p_j^{\ell m}] \cap [\ell^o])) \right| ds \underset{m \in M_j^{p^\ell}}{\cup} [p_{ej}^{\ell m}] ds \underset{m \in M_j^{p^\ell}}{\cup} [p_j^{\ell m}]/([p_j^{\ell m}] \cap [\ell^o])$$

Control of norm $\|.\|_{p_j^{\ell m},a}$ along the sequence of nodes $n_j^{p^\ell m}$ will be used to bound the $\|.\|_{1/\infty}$ norm, while control of norm $\|f(.,.,.)\|_{p_j^{\ell m},b}$ will be used to bound the $\|.\|_{\infty/1}$ norm. The idea behind these objects is that uniformity is relaxed to average case convergence with respect to those arguments for which uniform convergence is not required. These are the intermediate variables $s_{[p_{ej}^{\ell m}]}$ which will be integrated out by the final node and when controlling the $\|.\|_{1/\infty}$ error are output variables $t_{([p_j^{\ell m}] \cap [\ell^o])}$ in the final node and when controlling the $\|.\|_{\infty/1}$ error are input variables $s_{[p_j^{\ell m}]/([p_j^{\ell m}] \cap [\ell^o])}$. Bounds on these objects are aggregated over all incoming links $m \in M_j^{p^\ell}$ to form a bound on errors in $n_j^{p^\ell}$, which is defined using $\|f(.,.,.)\|_{p_j^{\ell m},a}$ for $\|.\|_{1/\infty}$ control and $\|f(.,.,.)\|_{p_j^{\ell m},b}$ for $\|.\|_{\infty/1}$ control. To ensure that these errors can be controlled, I consider a modified set of conditions which ensure appropriate bounds: replacing Condition (6), with a modified condition (6'):

Condition. (6') Same as Condition (6), but with the following changes

(i) Replace "Hölder continuous with exponent $\alpha_n^{p^\ell(p)}$ in input $n^p(.)$ (considered as a scalar) uniformly over all inputs other than p" with bounds

$$\left\| \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) - \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \right\|_{p_j^\ell,a}$$

$$\leq C \left(\sum_{p \in par(p)} \left\| \tilde{n}^{*p}(s_{[p]}) - n^{*p}(s_{[p]}) \right\|_\infty^{\alpha_d^{p^\ell(p)}} + c_{p^\ell(p)}^a \right)$$

$$\left\| \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) - \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \right\|_{p_j^\ell,b}$$

$$\leq C \left(\sum_{p \in par(p)} \left\| \tilde{n}^{*p}(s_{[p]}) - n^{*p}(s_{[p]}) \right\|_\infty^{\alpha_d^{p^\ell(p)}} + c_{p^\ell(p)}^b \right)$$

where $c_{p^\ell(p)}^a$ and $c_{p^\ell(p)}^b$ are constants depending on the node and which norm is being bounded such that $\min\{c_{p^\ell(p)}^a, c_{p^\ell(p)}^b\} = 0$

(v) Replace "sup norm accurate over classes $\hat{\mathcal{G}}_d^{[P_j^{\ell m}];[P_{ej}^{\ell m}]}$ at rate $\zeta^{\ell jm}(K_{[\chi+1]\times[j]})$" with "accurate over classes $\hat{\mathcal{G}}_d^{[P_j^{\ell m}];[P_{ej}^{\ell m}]}$ with respect to norm $\|.\|_{\infty/1}$ at rate $\zeta_b^{\ell jm}(K_{[\chi+1]\times[j]})$ and with respect to norm $\|.\|_{1/\infty}$ at rate $\zeta_a^{\ell jm}(K_{[\chi+1]\times[j]})$ "

(vi) Replace "sup norm accurate over classes $\hat{\mathcal{G}}_{dd}^{[P_j^{\ell m}];[P_{ej}^{\ell m}]}$ at rate $\zeta_d^{\ell jm}(K_{[\chi+1]\times[j]})$" with "accurate over classes $\hat{\mathcal{G}}_{dd}^{[P_j^{\ell m}];[P_{ej}^{\ell m}]}$ with respect to norm $\|.\|_{\infty/1}$ at rate $\zeta_{db}^{\ell jm}(K_{[\chi+1]\times[j]})$ and with respect to norm $\|.\|_{1/\infty}$ at rate $\zeta_{da}^{\ell jm}(K_{[\chi+1]\times[j]})$"

These conditions essentially weaken uniform conditions to conditions that may only be uniform in one dimension; for example, the first condition is still satisfied under Hölder continuity, but may also be valid for maps with forms like $1\{n(s) > s\}$ which may be discontinuous but may only have large error over a region of decreasing mass, resulting in error which decays for at least one of the norms. Replacing the uniform norm in the appropriate places with the more refined norms above, operator norm bounds for the accuracy of the solution can be achieved which take the same form, mutatis mutandis, as those based on uniform approximation of the functions. These are expressed in the following modified versions of Lemma (9) and Theorem (10).

Lemma. *(9') Let the model of interest satisfy conditions (17), (18), (19), and (20'). It is the case that* $\forall \ell$, $\forall p^\ell$, $\forall j = 1 \dots 2d_2 + o^\ell$, $\forall m \in M_j^{p^\ell}$

$$\left\| \tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')} \right\|_{p_j^{\ell m},a} \leq C \sum_{(p,\tilde{m}) \in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a)$$

$$\left\| \tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')} \right\|_{p_j^{\ell m},b} \leq C \sum_{(p,\tilde{m}) \in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^b)$$

implying that

$$\begin{aligned} \left\| \tilde{n}_j^{p^\ell} - n_j^{p^\ell} \right\|_{p_j^\ell,a} &\leq C \sum_{m \in M_j^{p^\ell}} \sum_{(p,\tilde{m}) \in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a) \\ \left\| \tilde{n}_j^{p^\ell} - n_j^{p^\ell} \right\|_{p_j^\ell,b} &\leq C \sum_{m \in M_j^{p^\ell}} \sum_{(p,\tilde{m}) \in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^b) \end{aligned} \tag{33}$$

where C is a constant not depending on K.

$\forall \ell = 1 \dots d_2$, if ℓ satisfies Condition (18)(b)(ii), $\forall j = 1 \dots 2d_2$

$$\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} (M_{[\ell^o]} \mathcal{F}^\ell_{\overrightarrow{g}_j(in)} (\Pi^{[j]})^{-1} M^*_{[j]})_{ik} \phi_i \phi_k - n_j^{P^\ell} \right\|_{1/\infty} \leq \epsilon_{Ka}^{(\ell,j)}$$

$$:= C \sum_{m \in M_j^{P^\ell}} (\zeta_a^{\ell jm}(K_{[\chi+1]\times[j]}) + \sum_{(p,\tilde{m}) \in dp(j,P^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a))$$

$$\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} (M_{[\ell^o]} \mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}} (\Pi^{[j]})^{-1} M^*_{[j]})_{ik} \phi_i \phi_k - n_j^{P^\ell} \right\|_{\infty/1} \leq \epsilon_{Kb}^{(\ell,j)}$$

$$:= C \sum_{m \in M_j^{P^\ell}} (\zeta_b^{\ell j m}(K_{[\chi+1]\times[j]}) + \sum_{(p,\tilde{m}) \in dp(j,P^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^b))$$

and if ℓ satisfies Condition (18)(b)(i), $\forall j = 1 \dots 2d_2$

$$\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} (M_{[\ell^o]} (\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^\ell(out))}})^{-1} \mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}} (\Pi^{[j]})^{-1} M^*_{[j]})_{ik} \phi_i \phi_k - n_j^{P^\ell} / n_{j(out)}^{P^\ell} \right\|_{1/\infty} \leq \epsilon_{Ka}^{(\ell,j)} :=$$

$$C(\sum_{m \in M_j^{P^\ell}} \zeta_{da}^{\ell j m}(K_{[\chi+1]\times[j]}) + \sum_{m \in M_j^{P^\ell}} \sum_{(p,\tilde{m}) \in dp(j,P^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a)$$
$$+ \sum_{m \in M_{j(out)}^{P^\ell}} \sum_{(p,\tilde{m}) \in dp(j(out),P^\ell,m)} (\zeta_{dK}^{pj(out)\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a))$$

$$\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} (M_{[\ell^o]} (\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^\ell(out))}})^{-1} \mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}} (\Pi^{[j]})^{-1} M^*_{[j]})_{ik} \phi_i \phi_k - n_j^{P^\ell} / n_{j(out)}^{P^\ell} \right\|_{\infty/1} \leq \epsilon_{Kb}^{(\ell,j)} :=$$

$$C(\sum_{m \in M_j^{P^\ell}} \zeta_{db}^{\ell j m}(K_{[\chi+1]\times[j]}) + \sum_{m \in M_j^{P^\ell}} \sum_{(p,\tilde{m}) \in dp(j,P^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^b)$$
$$+ \sum_{m \in M_{j(out)}^{P^\ell}} \sum_{(p,\tilde{m}) \in dp(j(out),P^\ell,m)} (\zeta_{dK}^{pj(out)\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^b))$$

where C are constants not depending on K.

Remark. For the purposes of this analysis, I have assumed that uniform approximation conditions for the steady state as guaranteed, for example, by Lemma (7), are still satisfied here even though that condition may not be satisfied for the model's derivatives. This is reasonable in cases in which, for example, the model equations are differentiable but with derivatives which are not continuous, as in the Huggett model with hard borrowing constraint, which implies a kinked policy function in steady state. In this case, uniformity may be achievable for the steady state but not for the derivatives of the model at the steady state. Analysis of the proof suggests that in the case where the steady state nodes converge at rate υ^p only in L^1 norm, the same bounds would apply except with $\alpha_d^{p(q)}$ replaced by $\min\{\alpha_d^{p(q)}, 1\}$. As it is not needed for the example, the full analysis, which would require a reformulation of Lemma (7), is not pursued here.

A modified version of Theorem (10) bounds the error in the operator norm implied by the algorithm in terms of the bounds from Lemma (A.3) and the more refined Young's inequality bound.

Theorem. *(10') Assume the conditions of Lemma (A.3). In addition assume the full system satisfies Condition (2) in Childers [2018]. Then Algorithm (1) produces approximate policy operators $\tilde{h}_x^K$, $\tilde{g}_x^K$ which satisfy the conditions of Childers [2018] Theorem 1 and so satisfies, for $\mathcal{H}^K = \underset{j\in 1\ldots d_x}{\times} Sp \otimes_{z\in[j]} \Phi_z$, and* $\epsilon_K = \max_{\ell\in 1\ldots d_2, j=1\ldots 2d_2} (\epsilon_{K,a}^{(\ell,j)})^{\frac{1}{2}} (\epsilon_{K,b}^{(\ell,j)})^{\frac{1}{2}}$

$$\sup_{\|f\|_{\mathcal{H}^K}=1} \left\| (\tilde{h}_x^K - h_x) f \right\| \leq O(\epsilon_K)$$

$$\sup_{\|f\|_{\mathcal{H}^K}=1} \left\| (\tilde{g}_x^K - g_x) f \right\| \leq O(\epsilon_K)$$

Proof of Theorem (10')

Proof. Identical to proof of Theorem (10) , except using the bounds from Lemma (9') and inequality (32) in place of the bound from Lemma (9) and the sup norm form of Young's inequality. □

Remark. In some cases, it will not be possible to ensure that $\epsilon_{Kb}^{(\ell,j)}$ and $\epsilon_{Ka}^{(\ell,j)}$ both converge to 0. Fortunately this is not needed here: so long as one converges and the other is bounded, one still obtains operator norm convergence. This is precisely the motivating case for this theorem, as in the presence of a discontinuity aligned with the axis of an input or output argument, uniform convergence over that argument cannot in general be achieved, but average convergence can. This permits either the $\|.\|_{1/\infty}$ norm of the error to converge or the $\|.\|_{\infty/1}$ norm, but not both.

Given this more refined bound, it is possible to demonstrate the consistency of the solution algorithm using certain classes of interpolation method even for models with discontinuities which do not permit uniform approximations. However, because uniformity is required along at least one dimension, special care is needed regarding the orientation of discontinuities when demonstrating convergence or lack thereof. As a result, membership in standard function classes, like Bounded Variation functions, piecewise regular functions (Mallat [2008] Ch 9), or their multivariate generalization, cartoon functions [Donoho, 2001], may not be sufficient provide control of convergence rates. Fortunately, modified versions of these classes may be defined which do ensure accurate interpolation with respect to $\|.\|_{1/\infty}$ or $\|.\|_{\infty/1}$ norm.

Below I define a class, modified from the class of cartoon functions defined in Definition 1 in Grohs et al. [2018], which allows discontinuities of the type encountered in practice and allows sharp approximation guarantees. These classes are motivated by models with borrowing or other constraints, which can induce lack of differentiability with respect to some arguments and discontinuity

with respect to others in derivatives, due to the possibility that constrained and unconstrained states induce discontinuously different responses. Given appropriate regularity conditions on the shape of these discontinuities and functions which are sufficiently regular elsewhere, local approximation methods like histograms and wavelets can confine the influence of these discontinuities to small enough sets to permit uniform and average approximation with respect to the appropriate arguments.

Definition 20. A function $f(x,y)$ $\mathcal{X} \times \mathcal{Y} \subseteq \mathbb{R}^{d_x} \times \mathbb{R}^{d_y}$ is in class $\mathcal{PR}_{1,0}^{\alpha,N}(\mathcal{X},\mathcal{Y})$ if it satisfies

(i) $f(x,.)$ is *uniformly piecewise regular in* y. Defining $N+1$ functions $\{f_n(.,.)\}_{n=0}^{N}$, and N sets $\{B_n\}_{n=1}^{N}$, I may write $f(.,.) = f_0(.,.) + \sum_{n=1}^{N} 1_{B_n} f_n(.,.)$, where $\{f_n(x,.)\}_{n=0}^{N}$ are uniformly $\alpha-$Hölder in y, i.e. $\forall n$, $\sup_x \|f_n(x,.)\|_{\Lambda^\alpha} < C$ and $B_n \subseteq \mathcal{Y}$ are compact domains with boundary ∂B_n which is a compact topologically embedded $\mathcal{C}^2$ hypersurface of $\mathbb{R}^{d_y}$ without boundary.

(ii) $\forall y \in \mathcal{Y}$, $f(.,y)$ is absolutely continuous on $\mathcal{X}$ and $\frac{\partial}{dx} f(.,y)$ is *uniformly piecewise regular in* x, where $\frac{\partial}{dx} f(.,y)$ is taken in the sense of a weak derivative.

I say $f(x,y) \in \mathcal{PR}_{0,1}^{\alpha,N}(\mathcal{X},\mathcal{Y})$ if the above conditions are satisfied with x and y reversed.

I say $f(x,y) \in \mathcal{PR}_{0,0}^{\alpha,N}(\mathcal{X},\mathcal{Y})$ if $f(x,.)$ is uniformly piecewise regular in y and $f(.,y)$ is uniformly piecewise regular in x.

Remark. These classes differ from the class of cartoon functions in a few notable ways. The major one is that the sets whose boundaries define the regions of discontinuities or kink points of the function are restricted to lie entirely in $\mathcal{X}$ or $\mathcal{Y}$, aligning them with the axes. This is not because discontinuities across these arguments cannot be handled, but because approximation accuracy in norms relevant to the method used here depends on the alignment of the discontinuity. I ask for at least one (weak) derivative with respect to those dimensions for which uniform convergence is needed, and so restrict discontinuities in that direction to the first derivative, while in directions for which average convergence is needed only, discontinuities of forms amenable to local approximation methods can be handled.[20] A minor difference is that a Hölder condition is imposed on the component functions rather than a continuous differentiability assumption: in the case where $d_y = 1$, this ensures that the marginal functions are piecewise regular in the sense of Mallat [2008] Theorem 9.12 and generally allows adapting the method of proof from that case.

[20] Handling discontinuities not aligned with axes presents no trouble. Given functions $\{f_n(.,.)\}_{n=N+1}^{N'}$ in $\Lambda^\alpha(\mathcal{X}\times\mathcal{Y})$ and sets $\{B_n\}_{n=N+1}^{N'}$ $B_n \subseteq \mathcal{X}\times\mathcal{Y}$ with boundary ∂B_n a $\mathcal{C}^2$ hypersurface in $\mathbb{R}^{d_x}\times\mathbb{R}^{d_y}$ such that $\max\{\sup_{x\in\mathcal{X}} \mu_x \partial B_n, \sup_{y\in\mathcal{Y}} \mu_y \partial B_n\} = 0$ where μ_x and μ_y are the marginal measure over $\mathcal{Y}$ conditioning on x and $\mathcal{X}$ conditioning on y, respectively, one could add $\sum_{n=N+1}^{N'} 1_{B_n} f_n(.,.)$ to $f \in \mathcal{PR}_{1,0}^{\alpha,N}(\mathcal{X},\mathcal{Y})$ or $f \in \mathcal{PR}_{0,1}^{\alpha,N}(\mathcal{X},\mathcal{Y})$ without affecting convergence rates of wavelet projection or thresholding approximations. Since this adds cumbersome notation and is not needed for the example problems, the matter is not pursued further.

Interpolation methods which ensure consistency over discontinuous functions are necessarily more restricted than those which do not. For polynomial interpolation at Chebyshev points of functions with finite sets of discontinuities, Campiti et al. [2012] show pointwise convergence at all points of continuity and uniform convergence over compact sets not containing discontinuities, but this does not seem to suffice to ensure norm convergence. In contrast, wavelet approximation is well known to ensure convergence in L^2 (and so also L^1) over bounded variation classes (see, e.g. Mallat [2008], Johnstone [2015] Ch. 9). Extending these results to $\|.\|_{1/\infty}$ convergence over $\mathcal{PR}_{0,1}^{\alpha,N}(\mathcal{X},\mathcal{Y})$ and $\|.\|_{\infty/1}$ convergence over $\mathcal{PR}_{1,0}^{\alpha,N}(\mathcal{X},\mathcal{Y})$ requires some modification, but is achievable in certain cases. Notably, it can be shown for histogram approximation, which is equivalent to a low order Daubechies wavelet interpolation. With some work, it can also be extended to higher order wavelets.

Lemma 21. *Let $d = d_x + d_y$ $\mathcal{Q} = \{\times_{z\in\{1,\dots d\}}\mathcal{T}_z, \otimes_{z\in\{1,\dots d\}}\Phi_z, \otimes_{z\in\{1,\dots d\}}M_z\}$ be the d^{th} order tensor product of identical histogram approximations of order K_z, with total cardinality $K = K_z^d$.*

(i) Let $\mathcal{F} = \mathcal{PR}_{1,0}^{\alpha,N}([0,1)^{d_x},[0,1)^{d_y})$ for $\alpha > \frac{1}{2}$. Then Q is accurate with respect to norm $\|.\|_{\infty/1}$ over class $\mathcal{F}$ at rate $\epsilon_K = O(K^{-\frac{\min\{\alpha,1\}}{d}})$, and with respect to norm $\|.\|_{1/\infty}$ over class $\mathcal{F}$ at rate $\epsilon_K = O(1)$ (i.e., error is bounded but need not converge).

(ii) Let $\mathcal{F} = \mathcal{PR}_{0,1}^{\alpha,N}([0,1)^{d_x},[0,1)^{d_y})$ for $\alpha > \frac{1}{2}$. Then Q is accurate with respect to norm $\|.\|_{\infty/1}$ over class $\mathcal{F}$ at rate $\epsilon_K = O(1)$, and with respect to norm $\|.\|_{1/\infty}$ over class $\mathcal{F}$ at rate $\epsilon_K = O(K^{-\frac{\min\{\alpha,1\}}{d}})$.

Remark. In the case where a final node $n_j^{P^\ell} \in \mathcal{PR}_{1,0}^{\alpha,N}([0,1),[0,1))$ or $\mathcal{PR}_{1,0}^{\alpha,N}([0,1),[0,1))$ with $\alpha \geq 1$, these rates, along with Theorem (A.3), imply that operator norm error of the solution algorithm is (at best) $O(K^{-\frac{1}{4}})$. This contrasts with a bound of $O(K^{-\frac{1}{2}})$ in the case of histograms for two-dimensional functions which are Lipschitz, by Lemma (17).

Alternate approach to discontinuities: restricted input classes

For functions with certain types of discontinuities, no known set of basis functions can ensure convergence of functional derivatives, even with respect to weaker norms which only require uniformity over a subset of input dimensions. A motivating example for this difficulty is the Euler equation in the Huggett model with hard borrowing constraints; because consumption today as a function of cash on hand today and consumption tomorrow as a function of cash on hand tomorrow both have derivatives which are discontinuous in their arguments, and both enter into the same equation, this kernel function of this equation cannot be approximated uniformly with respect to either using a known set of basis functions. Fortunately, this does not rule out that uniform approximation can be achieved using an *unknown* set of basis functions. Surprisingly, this

can be sufficient to ensure approximate solutions which converge to the truth, albeit over a more restricted class of inputs than previously discussed methods. The logic is that so long as the unknown basis can itself be approximated, one can compose the approximate solution with respect to the unknown basis, which converges due to uniformity, with an approximation of that basis, which converges at least over a limited class of inputs.

The idea behind constructing an unknown basis which ensures operator norm convergence of the functional derivatives of the equilibrium conditions and can itself be approximated with known functions is that the source of failure of uniform convergence is the presence of a discontinuity at a location which is only known approximately over a function which is piecewise smooth. If the location were known exactly, one could construct a set of basis functions with support covering the piecewise smooth regions which would produce an interpolation map which converges uniformly. With an approximately known location of the discontinuity, one can construct a representation approximately of this form, which differs from the unknown representation by a continuous deformation, which becomes closer and closer to an identity map as the approximation of the discontinuity becomes more accurate. Essentially, a regular histogram can be "stretched" or "squeezed" slightly so that the border of the bins aligns with the discontinuity. As a result, when applying the representation of the solution in terms of coefficients with respect to the unknown basis to the coefficients of a function with respect to the known basis, the approximation error can be controlled so long as the coefficients with respect to the basis functions which are used do not vary too much in response to small deformations. This property of small response to deformations is not one which applies to all possible bases or all possible input functions to which the solution may be applied, but it is possible to define such classes. These results are related to and inspired by the property of deformation invariance described in Bruna and Mallat [2013] and Grohs et al. [2018], though the use here applies a modified definition from the version described there in order to suit the features of the application.

To introduce these ideas, I introduce them concretely using a particular class of functions and interpolation schemes, based on histogram approximations adapted to the location of discontinuities.

Condition 22. Let $\mathcal{F} := \{f \in \mathcal{PR}_{0,0}^{\alpha,N}(\mathcal{X},\mathcal{Y}) : \{B_i\}_{i=1}^{N_x} = \{B_i^*\}_{i=1}^{N_x} \subseteq \mathcal{X}, \{B_i\}_{i=N_x+1}^{N_x+N_y} = \{B_i^*\}_{i=N_x+1}^{N_x+N_y} \subseteq \mathcal{Y}\}$ be a set of piecewise regular functions with a fixed set of regions over which the functions are piecewise smooth (corresponding to locations defined in steady state). Let $\mathcal{Q}_{d_x} = \{\times_{z\in\{1\ldots d_x\}}\mathcal{T}_z, \otimes_{z\in\{1\ldots d_x\}}\Phi_z, \otimes_{z\in\{1\ldots d_x\}}M_z\}$ and $\mathcal{Q}_{d_y} = \{\times_{z\in\{1\ldots d_y\}}\mathcal{T}_z, \otimes_{z\in\{1\ldots d_y\}}\Phi_z, \otimes_{z\in\{1\ldots d_y\}}M_z\}$ be tensor product histogram approximations of order $K_x := K_z^{d_x}$ and $K_y := K_z^{d_y}$ over $\mathcal{X}$ and $\mathcal{Y}$, respectively. Suppose there exist isomorphisms (continuous transforms with continuous inverses) $\tau_x^K : \mathcal{X} \to \mathcal{X}$ and $\tau_y^K : \mathcal{Y} \to \mathcal{Y}$ such that

(i) (a) $\forall i = 1 \ldots N_x \bigcup_{\{k:\ x_k \in B_i \cap \times_{z\in\{1\ldots d_x\}}\mathcal{T}_z\}} \{z = \tau_x^K(x) :\ x \in \text{support } \phi_k(.)\} = B_i$ and $\forall i = N_x+1 \ldots N_y \bigcup_{\{k:\ y_k \in B_i \cap \times_{z\in\{1\ldots d_y\}}\mathcal{T}_z\}} \{\tau_y^K(y) :\ y \in \text{support } \phi_k(.)\} =$

B_i (b) $\forall i = 1 \dots N_x$, $\{\partial B_i \cap \times_{z\in\{1\dots d_x\}} \mathcal{T}_z\} = \emptyset$ and $\forall i = N_x + 1 \dots N_x + N_y$, $\{\partial B_i \cap \times_{z\in\{1\dots d_y\}} \mathcal{T}_z\} = \emptyset$

(ii) $\max\{\left\|\tau_y^K(y) - y\right\|_\infty, \left\|\tau_x^K(x) - x\right\|_\infty\} \leq CK_z^{-1}$ for some universal constant C

(iii) (a) $\tau_x^K(x)$ and $\tau_y^K(y)$ and their inverses $\tau_x^{-1,K}(z)$ and $\tau_y^{-1,K}(w)$ are weakly differentiable, with Jacobians $\frac{\partial}{dz}\tau_x^{-1,K}(z)$ and $\frac{\partial}{dw}\tau_y^{-1,K}(w)$ such that $c_k^x := \int_{\{z\in\text{support }\phi_k(.)\}} \left(\det \frac{\partial}{dz}\tau_x^{-1,K}(\tau_x^K(z))\right)^{-1} dz$, $c_k^y := \int_{\{w\in\text{support }\phi_k(.)\}} \left(\det \frac{\partial}{dw}\tau_y^{-1,K}(\tau_y^K(w))\right)^{-1} dw$ satisfy $\frac{c_k^x}{1/K_z^{d_x}} - 1 = O(\frac{1}{K_z})$ and $\frac{c_k^y}{1/K_z^{d_y}} - 1 = O(\frac{1}{K_z})$

(b) Let $\left\|\left(\det \frac{\partial}{dz}\tau_x^{-1,K}(\tau_x^K(z))\right)^{-1}\right\|_{L^2(\mathcal{X})} \leq C$ and $\left\|\left(\det \frac{\partial}{dw}\tau_y^{-1,K}(\tau_y^K(w))\right)^{-1}\right\|_{L^2(\mathcal{Y})} \leq C$ for some constant independent of K.

Condition 22(i)(a) says that the transformations map the hypercube defining the support of each single histogram bin to a region contained entirely within the set B_i which contains the point whose value is interpolated to that bin, and that the set of these these transformed hypercubes forms a partition of the set B_i, for each region B_i. Effectively, this ensures that the transformed supports form a new partition of the sets $\mathcal{X}$ and $\mathcal{Y}$ with boundaries which exactly coincide with the locations of discontinuities, and that the new partition keeps the points to be interpolated on the same side of the boundary as their original location. Condition 22(i)(b) asks that the discontinuities are in "generic position" relative to the grid points, not intersecting them; as the discontinuities are of positive codimension, this will be satisfied for almost any choice of grid. It might fail, for example, if the discontinuities are located at rational points, but this can generally be solved by increasing K by an appropriate amount. Alternately, if the ratio is genuinely small, the discontinuity may be approximated exactly using finite regular grids and a transformation is not needed; as the transform is an analytical rather than computational device, this need not be known ex ante, and the stronger of the two sets of guarantees will hold.

This partition will be used to create a new (infeasible) interpolation scheme which permits uniform convergence. Conditions 22(ii) and (iii) impose additional regularity on the mapping creating this partition which ensures that the new bins remain close enough in location and volume, respectively, to the uniformly spaced rectangular bins, where the location measure is absolute and the volume measure is relative to the volume of a histogram bin. Condition 22 (iii)(a) bounds the total change in volume over each bin; condition 22 (iii)(b) bounds the norm of the distortion globally over the full space. Both are implied by the stronger conditions that $\left\|\det \frac{\partial}{dz}\tau_x^{-1,K}(\tau_x^K(z)) - 1\right\|_\infty$ and $\left\|\det \frac{\partial}{dw}\tau_y^{-1,K}(\tau_y^K(w)) - 1\right\|_\infty$ are each $O(\frac{1}{K_z})$. This latter condition is closer to the definition of a "small" deformation in Bruna and Mallat [2013] and Grohs et al. [2018], and may be convenient to verify in practice, but is not necessary for the results here.

While not needed for computations, and indeed not possible to compute, it is worth describing transformations which satisfy this condition, in order to demonstrate that such transformations exist and illustrate their properties.

Consider a function $f(x,y) \in \mathcal{PR}^{\alpha,4}_{0,0}([0,1],[0,1])$ with a discontinuity in x at a point c and likewise in y. For an evenly spaced histogram representation of order K, so long as condition 22(i)(b) holds, c will fall in some bin k^*, without loss of generality to the left of its center. I can construct a transform with the desired properties by defining the function

$$\tau^K(x) := \begin{cases} \frac{\frac{k^*-1}{K}}{c} x & x \leq \frac{k^*-1}{K} \\ \frac{k^*-1}{K} + (\frac{1-\frac{k^*-1}{K}}{1-c})(x - \frac{k^*-1}{K}) & x > \frac{k^*-1}{K} \end{cases}$$

This function satisfies $\left\|\tau^K_x(x) - x\right\|_\infty \leq \min\{c - \frac{k^*-1}{K}, \frac{k^*}{K} - c\} < \frac{1}{K}$, and has $c^x_k = \int_{\{x \in \text{support } \phi_k(.)\}} \left(\det \frac{\partial}{dz} \tau_x^{-1,K}(\tau^K_x(x))\right)^{-1} dx$ such that

$$\frac{c^x_k}{1/K} - 1 = \begin{cases} \frac{\frac{k^*-1}{K}}{c} - 1 & x_k \leq \frac{k^*-1}{K} \\ (\frac{1-\frac{k^*-1}{K}}{1-c}) - 1 & x_k > \frac{k^*-1}{K} \end{cases}$$

This can be seen to satisfy 22(ii) and (iii). Further, multidimensional tensor product versions of this construction can handle discontinuities which lie along straight lines with respect to each dimension. For discontinuities which do not lie along straight lines (as occurs for example in models with borrowing constraints in a single variable which depend on another, as in the version of the Huggett model with persistent income), somewhat more complicated structures may be needed. However, due to the limited volume of the boundaries of the discontinuities (see Grohs et al. [2018]), they will overlap at most a fraction $\frac{1}{K_z}$ of histogram bins, so, for large enough K, a transformation exists which moves each point no more than the diameter of a single bin, and spreads the change in volume over an entire region B_i, thus satisfying the displacement and volume conditions. For multidimensional piecewise regular functions with boundaries with non-zero curvature, transformations can be defined which shrink or stretch the edges of hypercubes intersecting the boundary towards or away from the interior of a smooth region by an amount less than the diameter of a hypercube and so of the specified order, though the formulas are less tractable.

Given a transformation that satisfies the above conditions, it is possible to construct an orthonormal basis with desirable approximation properties by forming histograms with bins defined by the transform of the support of the regularly spaced bins. Furthermore, the two histogram representations will be "close enough together" that only minimal approximation error is incurred if one uses the same coefficients in one as in the other.

Definition 23. Let $\mathcal{F} \subset \mathcal{PR}^{\alpha,N}_{0,0}(\mathcal{X},\mathcal{Y})$ be a set of functions, $\mathcal{Q}_{d_x}$, and $\mathcal{Q}_{d_y}$ be a set of histogram interpolation schemes and τ^K_x, τ^K_y be a set of transformations, all satisfying Condition 22. A *transformed histogram approximation* is a pair of interpolation schemes $\tilde{\mathcal{Q}}_{d_x} = \{\times_{z\in\{1\ldots d_x\}}\mathcal{T}_z, \tilde{\Phi}_x, \otimes_{z\in\{1\ldots d_x\}} M_z\}$ $\tilde{\mathcal{Q}}_{d_y} = \{\times_{z\in\{1\ldots d_y\}}\mathcal{T}_z, \tilde{\Phi}_y, \otimes_{z\in\{1\ldots d_y\}} M_z\}$ over $\mathcal{X}$ and $\mathcal{Y}$ respectively such that $\mathcal{T}_z$ and $M_z := \frac{1}{\sqrt{K_z}} I$ are exactly the same as in $\mathcal{Q}_{d_x}$, and $\mathcal{Q}_{d_y}$, and $\tilde{\Phi}_x = \{\tilde{\phi}_k(x)\}_{k=1}^{K_z^{d_x}}$,

$\tilde{\phi}_k(x) := \left(\frac{1}{c_k^x}\right)^{\frac{1}{2}} \mathbf{1}\{\tau_x^{-1,K}(x) \in \otimes_{z\in\{1\dots d_x\}}[\frac{k_z-1}{K}, \frac{k_z}{K})\}$ and $\tilde{\Phi}_y = \{\tilde{\phi}_k(y)\}_{k=1}^{K_z^{d_y}}$,
$\tilde{\phi}_k(y) := \left(\frac{1}{c_k^y}\right)^{\frac{1}{2}} \mathbf{1}\{\tau_y^{-1,K}(y) \in \otimes_{z\in\{1\dots d_y\}}[\frac{k_z-1}{K}, \frac{k_z}{K})\}$.

In words, this says that the transformed basis that is used retains the same evenly spaced interpolation points and evenly scaled interpolation maps as the standard histogram representation, but defines the bins over the transformation of the hypercubes. Note also that, in order to ensure that the basis functions represent an orthonormal basis, the indicator functions have been scaled by the square root of their volume.

A straightforward modification of the sup norm convergence results for histograms over Hölder continuous functions shows that a transformed histogram approximation is sup norm accurate over piecewise regular functions, with a slightly slower rate of convergence, arising from the fact that the unknown rescaling of the basis functions induces an error in the interpolation approximation which is of the order of the square root of the relative difference in volume of the original and transformed histogram bins. While this error is unfortunate, it is also incurred directly in the function representation for any input, so appears to be unavoidable, at least when using this class of approximations.[21]

Lemma 24. *Let $\mathcal{F} \subset \mathcal{PR}_{0,0}^{\alpha,N}(\mathcal{X}, \mathcal{Y})$ be a set of functions, $\mathcal{Q}_{d_x}$, and $\mathcal{Q}_{d_y}$ be a set of histogram interpolation schemes and τ_x^K, τ_y^K be a set of transformations, all satisfying Condition 22, and let $\tilde{\mathcal{Q}}_{d_x} \otimes \tilde{\mathcal{Q}}_{d_y}$ be a tensor product of transformed histogram approximations. Then $\tilde{\mathcal{Q}}_{d_x} \otimes \tilde{\mathcal{Q}}_{d_y}$ is sup norm accurate over $\mathcal{F}$ at rate $O(K_z^{-\min\{\alpha,\frac{1}{2}\}})$.*

Proof. See appendix. □

Using a transformed histogram representation, one can apply the algorithm exactly as described to a model defined in terms of functions with derivatives which are piecewise regular and generate a set of numerical solutions which converges to the truth at rates determined by the topological structure of the model and the sup norm rates for each subcomponent, by verifying the conditions of Lemma 9 or (9'). To see this, simply note that these algorithms never require evaluating the basis functions themselves, only the functions defining the equilibrium conditions of the model. The algorithm produces a set of matrices which represent a map defined over the coefficients of the transformed histogram representation of the input and output functions. However, unlike

[21] These results do suggest a possible way to improve the accuracy of the approximation beyond using more grid points; if one can apply numerical break finding schemes in steady state which locate the discontinuity with a higher order of accuracy than the distance between grid points, then the error from this term could be reduced. With a histogram approach, a method of locating discontinuities with error of order $\frac{1}{K_z^2}$ along with an adaptively chosen set of grid points could improve convergence rates to $O(K_z^{-\min\{\alpha,1\}})$, the same rate achievable with known discontinuity locations, albeit with convergence over a smaller class of functions. Whether further improvements could be possible using higher order wavelets with hard thresholding is an interesting avenue for future research.

for the other classes I have described, such a coefficient representation cannot be used directly, as the coefficients with respect to the transformed histogram basis of any function of interest are not known or calculable exactly.

Fortunately, a restricted set of functions can be approximated by simply using the coefficients with respect to a regular histogram representation; applying the reverse approximation to the output yields a mapping which is defined entirely in terms of computable quantities, and which inherits the accuracy guarantees accruing to the representation in terms of unknown basis functions. Precisely, define the following bound for any map defined in terms of the transformed histogram basis.

Lemma 25. *Assume the conditions of Lemma (24). Let* $Z^K \in \mathbb{R}^{K_x \times K_y}$ *be a matrix and* $Z \in \mathcal{L}(L^2(\mathcal{Y}) \to L^2(\mathcal{X}))$ *be an operator such that* $\left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} Z_{ij}^K \left\langle \tilde{\phi}_j(y), [.] \right\rangle \tilde{\phi}_i(x) - Z[.] \right\|_{op} \leq \epsilon_K \to 0$ *for some* ϵ_K*. Then for* Λ^{α_y} *the class of Hölder continuous functions in* $L^2(\mathcal{Y})$ *with Holder exponent* α_y*, for some constant* C*, obtain the bound.*

$$\sup_{\{f \in \Lambda^{\alpha_y} :\ \|f\|_{L^2(\mathcal{Y})}=1\}} \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} Z_{ij}^K \left\langle \phi_j(y), f(.) \right\rangle \phi_i(x) - Z[f(.)] \right\|_{L^2(\mathcal{X})} \leq \epsilon_K + C K_z^{-\min\{\alpha_y, \frac{1}{2}\}}$$

Proof. See appendix. □

The rate of convergence here reflects the error induced by representing functions with respect to the "wrong" basis. It is important that an L^2 norm be used, as transformation in general cannot ensure L^∞ convergence, which is precisely the reason for which the transformed basis is used to ensure operator norm convergence, which requires convergence of functions in a stronger sense. This lemma can be used to show convergence of representations of the solution operators $h_x,\ g_x$, and also impulse response functions and also derivatives of equilibrium conditions, if these need to be evaluated directly. More precisely

Corollary 26. *Let* $(\tilde{h}_x^K, \tilde{g}_x^K)$ *be a set of matrix representations of policy operators produced by applying Algorithm (1) using a tensor product histogram representation of order* K_z *in each dimension with bases* $\{\varphi_i(x) = \otimes_{z \in \{1 \dots d_x\}} \phi_{i_z}(s_z)\}_{i=1}^{K_z^{d_x}}$, $\{\varphi_i(x) = \otimes_{z \in \{1 \dots d_y\}} \phi_{i_z}(s_z)\}_{i=1}^{K_z^{d_y}}$ *to a model satisfying the conditions of Theorem (10) or (A.3) with convergence rate* ϵ_K *when using a transformed histogram approximation satisfying the conditions of Lemma (24). Then for* $\alpha \geq \frac{1}{2}$*, there exists* $\bar{K}$ *such that for any* $K > \bar{K}$

$$\sup_{\{f \in \Lambda^{\alpha} :\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[\tilde{h}_x^K\right]_{ij} \left\langle \varphi_j(x), f(.) \right\rangle \varphi_i(x) - h_x[f(.)] \right\|_{\mathcal{H}_X} \leq \epsilon_K + C K_z^{-\frac{1}{2}}$$

$$\sup_{\{f \in \Lambda^{\alpha} :\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[\tilde{g}_x^K\right]_{ij} \left\langle \varphi_j(x), f(.) \right\rangle \varphi_i(y) - g_x[f(.)] \right\|_{\mathcal{H}_Y} \leq \epsilon_K + C K_z^{-\frac{1}{2}}$$

Further, for any integer $m \geq 1$, for any $K > \bar{K}$, the numerical impulse response functions derived from above solutions satisfy

$$\sup_{\{f \in \Lambda^\alpha:\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[(\tilde{h}_x^K)^m \right]_{ij} \langle \varphi_j(x), f(.) \rangle \, \varphi_i(x) - (h_x)^m [f(.)] \right\|_{\mathcal{H}_X} \leq C(\epsilon_K + K_z^{-\frac{1}{2}})$$

$$\sup_{\{f \in \Lambda^\alpha:\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[\tilde{g}_x^K (\tilde{h}_x^K)^m \right]_{ij} \langle \varphi_j(x), f(.) \rangle \, \varphi_i(y) - g_x(h_x)^m [f(.)] \right\|_{\mathcal{H}_Y} \leq C(\epsilon_K + K_z^{-\frac{1}{2}})$$

where the constant factor does not depend on K or m, i.e., the convergence is uniform over the length of the impulse response.

Proof. See appendix. □

Remark. The first claim follows by verification of conditions of Lemma (25), plus a bound on the direct error from using a transformed histogram representation of f. The second follows from the first and repeated application of the triangle inequality, where uniformity holds because impulse responses to stable models decay geometrically to 0. A separate convergence result for impulse response functions is needed because the output of a policy operator applied to a Hölder smooth function need not itself be Hölder smooth.

Under mild conditions on the functions defining the equilibrium conditions of the model and its derivatives, such as that they form functions in spaces $\mathcal{PR}_{0,0}^{\alpha,N}$ for $\alpha \geq \frac{1}{2}$ and that each node is Hölder continuous with respect to parent nodes with exponent at least 1, the term ϵ_K in the above corollary will also be of order $K_z^{-\frac{1}{2}}$, making this the rate for the model as a whole. For example, if, as in the Huggett models with i.i.d. income, the kernel functions defining the model are $2-$dimensional, an order K histogram representation will have a convergence rate of $O(K^{-\frac{1}{4}})$. For versions of these models where income is persistent and so the dimension rises to 4 and so $K = K_z^4$, this rate slows to $O(K^{-\frac{1}{8}})$. This compares to rates of $O(K^{-\frac{1}{2}})$ and $O(K^{-\frac{1}{4}})$ for the infeasible version of a histogram representation, when the location of any discontinuity is known. In this infeasible case, when the regularity α of the continuous components of the functions is higher than $\frac{1}{2}$, even faster rates are possible using bases other than histograms. These improvements suggest at least two possible avenues for improvement: use of methods which locate discontinuities more precisely, and use of bases adapted to higher levels of regularity. As rates for the current method may be prohibitively slow in some applications requiring many evaluations or high precision, these represent important practical directions for future work.

B Additional Technical Results

The equivalence between certain methods based on discretization followed by differentiation and others based on differentiation then discretization arises due to conditions which ensure the equivalency of different approaches in terms of the resulting solution. In this appendix, I present a set of generic high level equivalency results, which may then be used to show equivalency of particular methods.

Lemma 27. *Let $F(.): \mathcal{H}_1 \to \mathcal{H}_2$ be an operator between two separable Hilbert spaces, equipped with complete orthonormal bases $\{\phi_i\}_{i=1}^{\infty}$ and $\{\varphi_j\}_{j=1}^{\infty}$ respectively, and let π^{K_1} and π^{K_2} represent orthogonal projection onto the span of $\{\phi_i\}_{i=1}^{K_1}$ and $\{\varphi_j\}_{j=1}^{K_2}$ respectively. Assume F is continuously[22] Fréchet differentiable in a neighborhood of $g^* \in \mathcal{H}_1$, with derivative of F at g^* given by $F_{g^*}[.]$. Then $\frac{d}{d\langle g,\phi_i\rangle}\left\langle F(\sum_{s=1}^{K_1}\langle g^*,\phi_s\rangle\phi_s),\varphi_j\right\rangle \to \langle F_{g^*}[\phi_i],\varphi_j\rangle$, i.e. the scalar derivative of the j^{th} basis coefficient of the output function with respect to the i^{th} basis coefficient of the projection π^{K_1} of the input function, evaluated at g^*, converges to the $(i,j)^{th}$ entry of the matrix representation of the functional derivative $\pi^{K_2}F_{g^*}\pi^{K_1}$ as $K_1 \to \infty$.*

Proof.

$$\begin{aligned}
\frac{d}{d\langle g,\phi_i\rangle}\left\langle F(\sum_{s=1}^{K_1}\langle g^*,\phi_s\rangle\phi_s),\varphi_j\right\rangle &= \left\langle \frac{d}{d\langle g,\phi_i\rangle}F(\sum_{s=1}^{K_1}\langle g^*,\phi_s\rangle\phi_s),\varphi_j\right\rangle \\
&= \left\langle F_g(\sum_{s=1}^{K_1}\langle g^*,\phi_s\rangle\phi_s)[\phi_i],\varphi_j\right\rangle \\
&\to \langle F_{g^*}[\phi_i],\varphi_j\rangle
\end{aligned}$$

where the first line holds by linearity of inner products and boundedness of the functional derivatives, the second line holds by the chain rule for Fréchet derivatives, and the final line by the assumption of continuous differentiability and the fact that $\left\|g^* - \pi^{K_1}g^*\right\|_{\mathcal{H}_1} \to 0$ by completeness of the basis. □

In words, what this result says is that rather than calculating functional derivatives and then projecting them onto the span of a set of basis functions, I can instead first represent the inputs and outputs in terms of their basis function coefficients and then take the derivatives of the output coefficients, which are scalars, with respect to the input coefficients, which are also scalars, and produce the same matrix, at least asymptotically. The assumption of continuous differentiability and the asymptotic nature of the result arise only from the fact that the derivative need not be taken at the exact steady state, but can instead be taken at an approximation thereof. This is not truly unique to the scalar derivative case, and such an assumption would also be needed

[22] Continuity is assumed to hold on the space of functions $\mathcal{H}_1 \to \mathcal{L}(\mathcal{H}_1 \to \mathcal{H}_2)$: i.e., convergence in $\mathcal{H}_1$ implies convergence in operator norm of the derivative.

even with directly calculated functional derivatives whenever the steady state is approximated numerically.

Constructing projections of functional derivatives in this way allows replacing steps 2 and 4 in Algorithm 1 of Childers [2018] with a single automated procedure. When used in this way, one must still be sure that the conditions for validity of this algorithm hold: in particular, it must be the case that the functional derivatives satisfy the condition of separability into identity and compact components. This separation is also necessary for the construction of the analytical correction term derived from the identity component. When automating the model linearization and approximation process, there are two ways to handle this term. First, one can simply ignore it, and construct only the approximation using the projected derivatives, calculated automatically. While the solution generated by such a process is no longer consistent in uniform norm, it is consistent over the class of inputs which lie in the span of the projected space: for example, for an approximation based on K polynomials, the error bound in Childers [2018] Theorem 1 applies uniformly over all input functions which are K^{th} order polynomials of norm 1 or less. Second, to construct the correction term, it suffices to construct only the identity component of the functional derivatives, which can be done without analytical knowledge of the other components of the functional derivatives, which may be handled completely automatically. Identifying the identity component is in general not challenging, and I will provide a scheme for doing so in a particular but fairly general class of models to which the procedure applies. Once it is identified, the correction term can be constructed in the usual manner without needing any knowledge of the functional derivatives of the other components.

While the above result permits the use of standard scalar differentiation methods to produce the projected functional derivatives used in the solution algorithm of Childers [2018], it is only useful in practice when the action of the nonlinear operator on the input basis functions and the basis coefficients of the output function can be calculated or approximated in a way which ensures that each scalar function has derivatives which can be calculated by standard automatic differentiation software. In most cases of practical interest, as inner products define integrals which rarely have simple closed form representations, these functions must be approximated numerically.

Additional Invariances

Here I provide a useful generic invariance result for the solutions provided by a rational expectations algorithm. The solution is invariant to left multiplication by invertible transformations of the matrices representing the derivatives of the equilibrium conditions. In other words, a change of basis leads to the same solution, expressed in the new basis. For right multiplication, the solutions differ, but in a way which can be resolved by multiplying the solutions by the appropriate inverse transformations; in other words, a change of basis for the derivatives produces only the equivalent change of basis for the solutions, and no other change.

Lemma 28. *Transformation invariance. Let* (B, A) *be a* $(K_x+K_y)\times(K_x+K_y)$ Γ*-regular matrix pencil with generalized Schur decomposition with respect to closed curve* Γ *given by* (Q^*SU, Q^*TU), $(S,T) = \left(\begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix}, \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix}\right)$, $U = \begin{bmatrix} U_{11} & U_{12} \\ U_{21} & U_{22} \end{bmatrix}$ *resulting in standard (approximate) rational expectation solution operators* $g_x = -U_{22}^{-1}U_{21}$, $h_x = (U_{11}+U_{12}g_x)^{-1}S_{11}^{-1}T_{11}(U_{11}+U_{12}g_x)$. *Let* W^l *and* $W^r = \begin{bmatrix} W_x^r & 0 \\ 0 & W_y^r \end{bmatrix}$ *be* $(K_x+K_y)\times(K_x+K_y)$ *matrices representing transformations of* (B, A)*, and suppose* W^l *and* W^r *are invertible. Then the solution operators corresponding to the matrix pencil* (W^lBW^r, W^lAW^r) *are given by* $g_x^r = W_y^{r-1}g_xW_x^r$ *and* $h_x^r = W_x^{r-1}h_xW_x^r$*, respectively.*

Proof. Follows from straightforward linear algebra. See Appendix. □

This lemma is useful whenever a particular choice of representation has desirable computational properties. In particular, when a basis can be (approximately) represented in terms of pointwise evaluations, it permits the use of these evaluations instead of the basis function coefficients as the state variables in the solution algorithm, and so, with appropriate transformations ex post, permits use of pointwise derivatives rather than functional derivatives.

B.1 Functional Derivatives of Canonical Heterogeneous Agent Model

Recall that predetermined state variables are $Lf_{X,t}(X_{it}), Lg_t(X_{it},\epsilon_{it}), LP_t, P_{2t}$, and jump variables are $f_{X,t}(X_{it})$, $g_t(X_{it},\epsilon_{it})$, P_{1t} and $V_t(X_{it},\epsilon_{it})$.

The derivatives of the equilibrium conditions are as follows.

Equation (6)

$\frac{d}{dLf_X}F:\ 0$

$\frac{d}{dLg}F:\ 0$

$\frac{d}{dLP}F:\ 0$

$\frac{d}{dP_2}F:\ \frac{\partial}{dP_{2t}}u(g^*(X_{it},\epsilon_{it}),X_{it},P^*,\epsilon_{it})\cdot[.]+$

$+\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\frac{\partial}{dP_{2t}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]$

$f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$

$+\beta\int\int V^*(X_{it+1},\epsilon_{it+1})f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))f_\epsilon(\epsilon_{it+1},P^*)\cdot$

$\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|\left(\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1}$

$\frac{\partial^2}{dP_{2t}X_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]dX_{it+1}d\epsilon_{it+1}$

$\frac{d}{df_X}F:\ 0$

$\frac{d}{dg}F:\ u_Y(g^*(X_{it},\epsilon_{it}),X_{it},P^*,\epsilon_{it})\cdot[.]+$

$\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\frac{\partial}{dg}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]$

$f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$

$+\beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) f_\epsilon(\epsilon_{it+1}, P^*) \cdot$
$\left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| \left(\frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1}$
$\frac{\partial^2}{dgX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.] dX_{it+1} d\epsilon_{it+1}$
$\frac{d}{dP_1} F: \ \frac{\partial}{dP_{1t}} u(g^*(X_{it}, \epsilon_{it}), X_{it}, P^*, \epsilon_{it}) \cdot [.] +$
$+\beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) \frac{\partial}{dU} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) \frac{\partial}{dP_{1t}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]$
$f_\epsilon(\epsilon_{it+1}, P^*) \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1} d\epsilon_{it+1}$
$+\beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) f_\epsilon(\epsilon_{it+1}, P^*) \cdot$
$\left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| \left(\frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1}$
$\frac{\partial^2}{dP_{1t} X_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.] dX_{it+1} d\epsilon_{it+1}$
$\frac{d}{dV} F: \ -I$
$\frac{d}{dLf'_X} F: \ 0$
$\frac{d}{dLg'} F: \ 0$
$\frac{d}{dLP'} F: \ 0$
$\frac{d}{dP'_2} F: \beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))$
$[\frac{\partial}{dP_{2t+1}} f_\epsilon(\epsilon_{it+1}, P^*)][.] \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1} d\epsilon_{it+1}$
$+\beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) \frac{\partial}{dU} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) \frac{\partial}{dP_{2t+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]$
$f_\epsilon(\epsilon_{it+1}, P^*) \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1} d\epsilon_{it+1}$
$+\beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) f_\epsilon(\epsilon_{it+1}, P^*) \cdot$
$\left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| \left(\frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1}$
$\frac{\partial^2}{dP_{2t+1} X_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.] dX_{it+1} d\epsilon_{it+1}$
$\frac{d}{df'_X} F: \ 0$
$\frac{d}{dg'} F: \ 0$
$\frac{d}{dP'_1} F: \beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))$
$[\frac{\partial}{dP_{1t+1}} f_\epsilon(\epsilon_{it+1}, P^*)][.] \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1} d\epsilon_{it+1}$
$+\beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) \frac{\partial}{dU} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) \frac{\partial}{dP_{1t+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]$
$f_\epsilon(\epsilon_{it+1}, P^*) \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1} d\epsilon_{it+1}$
$+\beta \int \int V^*(X_{it+1}, \epsilon_{it+1}) f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) f_\epsilon(\epsilon_{it+1}, P^*) \cdot$
$\left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| \left(\frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1}$
$\frac{\partial^2}{dP_{1t+1} X_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.] dX_{it+1} d\epsilon_{it+1}$
$\frac{d}{dV'} F: \ \beta \int \int [.] f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) f_\epsilon(\epsilon_{it+1}, P^*) \cdot$
$\left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1} d\epsilon_{it+1}$
Equation (7)
$\frac{d}{dLf_X} F: \ 0$

$\frac{d}{dLg}F:\ 0$
$\frac{d}{dLP}F:\ 0$
$\frac{d}{dP_2}F:\ u_{YP_2}(g^*(X_{it},\epsilon_{it}),X_{it},P^*,\epsilon_{it})\cdot[.]+$
$\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial^2}{dU^2}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\frac{\partial}{dP_{2t}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]\cdot$
$\frac{\partial}{dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})f_\epsilon(\epsilon_{it+1},P^*,P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$
$+\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU_{it+1}}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial^2}{dP_{2t}dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$
$+\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU_{it+1}}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial}{dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|$
$\left(\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1}\frac{\partial^2}{dP_{2t}dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]dX_{it+1}d\epsilon_{it+1}$
$\frac{d}{df_X}F:\ 0$
$\frac{d}{dg}F:\ u_{YY}(g^*(X_{it},\epsilon_{it}),X_{it},P^*,\epsilon_{it})\cdot[.]+$
$\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial^2}{dU^2}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\frac{\partial}{dg}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]\cdot$
$\frac{\partial}{dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$
$+\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU_{it+1}}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial^2}{dYdY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$
$+\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU_{it+1}}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial}{dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|$
$\left(\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1}\frac{\partial^2}{dYdX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]dX_{it+1}d\epsilon_{it+1}$
$\frac{d}{dP_1}F:\ u_{YP_1}(g^*(X_{it},\epsilon_{it}),X_{it},P^*,\epsilon_{it})\cdot[.]+$
$\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial^2}{dU^2}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\frac{\partial}{dP_{1t}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]\cdot$
$\frac{\partial}{dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$
$+\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU_{it+1}}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial^2}{dP_{1t}dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$
$+\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial}{dU_{it+1}}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial}{dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|$
$\left(\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1}\frac{\partial^2}{dP_{1t}dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]dX_{it+1}d\epsilon_{it+1}$
$\frac{d}{dV}F:\ 0$
$\frac{d}{dLf'_X}F:\ 0$
$\frac{d}{dLg'}F:\ 0$
$\frac{d}{dLP'}F:\ 0$
$\frac{d}{dP'_2}F:\beta\int\int V^*(X_{it+1},\epsilon_{it+1})\frac{\partial^2}{dU^2}f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\frac{\partial}{dP_{2t+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]\cdot$
$\frac{\partial}{dY}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})f_\epsilon(\epsilon_{it+1},P^*)\left|\det\frac{\partial}{dX_{it+1}}Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|dX_{it+1}d\epsilon_{it+1}$

$+\beta \int\int V^*(X_{it+1},\epsilon_{it+1}) \frac{\partial}{dU_{it+1}} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial^2}{dP_{2t+1}dY} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.] f_\epsilon(\epsilon_{it+1},P^*) \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1}d\epsilon_{it+1}$
$+\beta \int\int V^*(X_{it+1},\epsilon_{it+1}) \frac{\partial}{dU_{it+1}} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial}{dY} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}) \frac{\partial}{dP_{2t+1}} f_\epsilon(\epsilon_{it+1},P^*)[.] \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1}d\epsilon_{it+1}$
$+\beta \int\int V^*(X_{it+1},\epsilon_{it+1}) \frac{\partial}{dU_{it+1}} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial}{dY} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}) f_\epsilon(\epsilon_{it+1},P^*) \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|$
$\left(\frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1} \frac{\partial^2}{dP_{2t+1}dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.] dX_{it+1}d\epsilon_{it+1}$
$\frac{d}{df'_X} F:\ 0$
$\frac{d}{dg'} F:\ 0$
$\frac{d}{dP'_1} F: \beta \int\int V^*(X_{it+1},\epsilon_{it+1}) \frac{\partial^2}{dU^2} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})) \frac{\partial}{dP_{1t+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.]\cdot$
$\frac{\partial}{dY} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}) f_\epsilon(\epsilon_{it+1},P^*) \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1}d\epsilon_{it+1}$
$+\beta \int\int V^*(X_{it+1},\epsilon_{it+1}) \frac{\partial}{dU_{it+1}} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial^2}{dP_{1t+1}dY} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.] f_\epsilon(\epsilon_{it+1},P^*) \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1}d\epsilon_{it+1}$
$+\beta \int\int V^*(X_{it+1},\epsilon_{it+1}) \frac{\partial}{dU_{it+1}} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial}{dY} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}) \frac{\partial}{dP_{1t+1}} f_\epsilon(\epsilon_{it+1},P^*)[.] \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1}d\epsilon_{it+1}$
$+\beta \int\int V^*(X_{it+1},\epsilon_{it+1}) \frac{\partial}{dU_{it+1}} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot$
$\frac{\partial}{dY} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}) f_\epsilon(\epsilon_{it+1},P^*) \left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right|$
$\left(\frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right)^{-1} \frac{\partial^2}{dP_{1t+1}dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})[.] dX_{it+1}d\epsilon_{it+1}$
$\frac{d}{dV'} F:\ \beta \int\int [.] \frac{\partial}{dU_{it+1}} f_U(Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}))\cdot \frac{\partial}{dY} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1}) f_\epsilon(\epsilon_{it+1},P^*)\cdot$
$\left|\det \frac{\partial}{dX_{it+1}} Q^{-1}_{g^*(X_{it},\epsilon_{it}),P^*,P^*,X}(X_{it+1})\right| dX_{it+1}d\epsilon_{it+1}$

Equation (10)

$\frac{d}{dLf_X} F:\ 0$
$\frac{d}{dLg} F:\ 0$
$\frac{d}{dLP} F:\ 0$
$\frac{d}{dP_2} F:\ \frac{\partial}{dP_2} F(f^*_X(X_{it}), g^*(X_{it},\epsilon_{it}), P^*)[.]$
$\frac{d}{df_X} F:\ \frac{\partial}{df_X} F(f^*_X(X_{it}), g^*(X_{it},\epsilon_{it}), P^*)[.]$
$\frac{d}{dg} F:\ \frac{\partial}{dg} F(f^*_X(X_{it}), g^*(X_{it},\epsilon_{it}), P^*)[.]$
$\frac{d}{dP_1} F:\ \frac{\partial}{dP_1} F(f^*_X(X_{it}), g^*(X_{it},\epsilon_{it}), P^*)[.]$
$\frac{d}{dV} F:\ 0$
$\frac{d}{dLf'_X} F:\ 0$
$\frac{d}{dLg'} F:\ 0$
$\frac{d}{dLP'} F:\ 0$
$\frac{d}{dP'_2} F:\ 0$
$\frac{d}{df'_X} F:\ 0$

$\frac{d}{dg'}F:\ 0$
$\frac{d}{dP_1'}F:\ 0$
$\frac{d}{dV'}F:\ 0$
Equation (11)
$\frac{d}{dLf_X}F:\ 0$
$\frac{d}{dLg}F:\ 0$
$\frac{d}{dLP}F:\ 0$
$\frac{d}{dP_2}F:\ \frac{d}{dP_2}h(P_2^*,0)$
$\frac{d}{df_X}F:\ 0$
$\frac{d}{dg}F:\ 0$
$\frac{d}{dP_1}F:\ 0$
$\frac{d}{dV}F:\ 0$
$\frac{d}{dLf_X'}F:\ 0$
$\frac{d}{dLg'}F:\ 0$
$\frac{d}{dLP'}F:\ 0$
$\frac{d}{dP_2'}F:\ -I$
$\frac{d}{df_X'}F:\ 0$
$\frac{d}{dg'}F:\ 0$
$\frac{d}{dP_1'}F:\ 0$
$\frac{d}{dV'}F:\ 0$
Equation (13)(i)
$\frac{d}{dLf_X}F:\ 0$
$\frac{d}{dLg}F:\ 0$
$\frac{d}{dLP}F:\ 0$
$\frac{d}{dP_2}F:\ 0$
$\frac{d}{df_X}F:\ -I$
$\frac{d}{dg}F:\ 0$
$\frac{d}{dP_1}F:\ 0$
$\frac{d}{dV}F:\ 0$
$\frac{d}{dLf_X'}F:\ I$
$\frac{d}{dLg'}F:\ 0$
$\frac{d}{dLP'}F:\ 0$
$\frac{d}{dP_2'}F:\ 0$
$\frac{d}{df_X'}F:\ 0$
$\frac{d}{dg'}F:\ 0$
$\frac{d}{dP_1'}F:\ 0$
$\frac{d}{dV'}F:\ 0$
Equation (13)(ii)
$\frac{d}{dLf_X}F:\ 0$

$\frac{d}{dLg}F:\ 0$
$\frac{d}{dLP}F:\ 0$
$\frac{d}{dP_2}F:\ 0$
$\frac{d}{df_X}F:\ 0$
$\frac{d}{dg}F:\ -I$
$\frac{d}{dP_1}F:\ 0$
$\frac{d}{dV}F:\ 0$
$\frac{d}{dLf'_X}F:\ 0$
$\frac{d}{dLg'}F:\ I$
$\frac{d}{dLP'}F:\ 0$
$\frac{d}{dP'_2}F:\ 0$
$\frac{d}{df'_X}F:\ 0$
$\frac{d}{dg'}F:\ 0$
$\frac{d}{dP'_1}F:\ 0$
$\frac{d}{dV'}F:\ 0$
Equation (13)(iii)
$\frac{d}{dLf_X}F:\ 0$
$\frac{d}{dLg}F:\ 0$
$\frac{d}{dLP}F:\ 0$
$\frac{d}{dP_2}F:\ -I$
$\frac{d}{df_X}F:\ 0$
$\frac{d}{dg}F:\ 0$
$\frac{d}{dP_1}F:\ -I$
$\frac{d}{dV}F:\ 0$
$\frac{d}{dLf'_X}F:\ 0$
$\frac{d}{dLg'}F:\ 0$
$\frac{d}{dLP'}F:\ I$
$\frac{d}{dP'_2}F:\ 0$
$\frac{d}{df'_X}F:\ 0$
$\frac{d}{dg'}F:\ 0$
$\frac{d}{dP'_1}F:\ 0$
$\frac{d}{dV'}F:\ 0$
Equation (15)
$\frac{d}{dLf_X}F:\ \int\int f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\cdot$
$\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|[.]f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$
$\frac{d}{dLg}F:\ \int\int \frac{\partial}{dU}f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\frac{\partial}{dY}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it-1})[.]$
$\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|Lf^*_X(X_{it-1})f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$

$+\int\int f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|$
$\left(\frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right)^{-1}\frac{\partial^2}{dYdX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})[.]$
$Lf^*_X(X_{it-1})f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$
$\frac{d}{dLP}F:\ \int\int \frac{\partial}{dU}f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\frac{\partial}{dLP}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})[.]$
$\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|Lf^*_X(X_{it-1})f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$
$+\int\int f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|$
$\left(\frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right)^{-1}\frac{\partial^2}{dLPdX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})[.]$
$Lf^*_X(X_{it-1})f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$
$+\int\int f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|$
$Lf^*_X(X_{it-1})\frac{\partial}{dLP}f_\epsilon(\epsilon_{it-1},LP^*)[.]dX_{it-1}d\epsilon_{it-1}$
$\frac{d}{dP_2}F:\ \int\int \frac{\partial}{dU}f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\frac{\partial}{dP_2}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})[.]$
$\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|Lf^*_X(X_{it-1})f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$
$+\int\int f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|$
$\left(\frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right)^{-1}\frac{\partial^2}{dP_2dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})[.]$
$Lf^*_X(X_{it-1})f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$
$\frac{d}{df_X}F:\ -I$
$\frac{d}{dg}F:\ 0$
$\frac{d}{dP_1}F:\ \int\int \frac{\partial}{dU}f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\frac{\partial}{dP_1}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})[.]$
$\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|Lf^*_X(X_{it-1})f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$
$+\int\int f_U(Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it}))\left|\det \frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right|$
$\left(\frac{\partial}{dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})\right)^{-1}\frac{\partial^2}{dP_1dX_{it}}Q^{-1}_{Lg^*(X_{it-1},\epsilon_{it-1}),LP^*,P^*,X_{it-1}}(X_{it})[.]$
$Lf^*_X(X_{it-1})f_\epsilon(\epsilon_{it-1},LP^*)dX_{it-1}d\epsilon_{it-1}$
$\frac{d}{dV}F:\ 0$
$\frac{d}{dLf'_X}F:\ 0$
$\frac{d}{dLg'}F:\ 0$
$\frac{d}{dLP'}F:\ 0$
$\frac{d}{dP'_2}F:\ 0$
$\frac{d}{df'_X}F:\ 0$
$\frac{d}{dg'}F:\ 0$
$\frac{d}{dP'_1}F:\ 0$
$\frac{d}{dV'}F:\ 0$

C Proofs

C.1 Proofs of Claims from Main Text

Proof. of Lemma (7)

Statement (i) follows from 5(iii) and (ii), and the fact that each node is defined iteratively by composing the listed operators. For each ℓ, inequality (27) in statement (ii) can be shown by induction along the graph, from input nodes $1 \dots 2d_2 + o_\ell$ to output node P^ℓ. Input nodes $\tilde{n}^{*p^\ell}$ for $p^\ell = 1 \dots 2d_2 + o_\ell$ are given by $\tilde{n}^{*p^\ell} = \tilde{g}^*_{j(p^\ell)}(s_{[j(p^\ell)]}) \in \hat{\mathcal{G}}^{[p^\ell]} = \hat{\mathcal{G}}^{[p^\ell_a] \cup [p^\ell_n]}$ by assumption 5(iv) and (v), and so by 5 (v), $\left\| \tilde{n}^{*p^\ell} - n^{*p^\ell} \right\|_\infty \leq \zeta^{p^\ell}(K_{[p^\ell_a] \cup [p^{\ell}{}_n]})$, which shows the inductive base case. Now let $p^\ell \in 1 \dots P^\ell$ be such that all ancestor nodes s of p^ℓ satisfy $\|\tilde{n}^{*s} - n^{*s}\|_\infty \leq C_s(\zeta^{[s_b]}(K_{[s_b]})1_s + \sum_{j \in dp(p^\ell)} \sum_{p \in j} (\zeta^{[p_b(e)]}(K_{[p_b(e)]})1_p)^{\alpha^{p,j}})$ and $\tilde{n}^{*s} \in \hat{\mathcal{G}}^{[s]}$. Then, in the case where node p^ℓ has identity for a linear component, obtain bound $\left\| \tilde{n}^{*p^\ell} - n^{*p^\ell} \right\|_\infty$

$$
\begin{aligned}
&= \left\| f^{p^\ell}(s_{[p^\ell_I]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) - f^{p^\ell}(s_{[p^\ell_I]}, \{n^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \right\|_\infty \\
&\leq C \sum_{p \in par(p^\ell)} \left\| \tilde{n}^{*p}(s_{[p]}) - n^{*p}(s_{[p]}) \right\|_\infty^{\alpha^{p^\ell(p)}} \\
&\leq 0 \cdot \zeta^s(K_s) + C' \sum_{j \in dp(p^\ell)} \sum_{p \in j} (\zeta^{p(e)}(K_{p(e)})1_p)^{\alpha^{p,j}})
\end{aligned}
$$

The first line follows by definition of $\tilde{n}^{*p^\ell}$ and n^{*p^ℓ}, and the second by Condition 5(i). The last line follows from the inductive hypothesis and the inequality that $(a+b)^\alpha \leq C_\alpha(a^\alpha + b^\alpha)$ for some $C_\alpha > 0$, for all $a,\ b \geq 0$ (C_α may be taken as 1 for $\alpha \leq 1$). $\tilde{n}^{*p^\ell} \in \hat{\mathcal{G}}^{[p^\ell]}$ by Conditions 5(iv) (a) and (b).

In the case where node p^ℓ has an integral operator for a linear component, obtain bound $\left\| \tilde{n}^{*p^\ell} - n^{*p^\ell} \right\|_\infty$

$$
\begin{aligned}
=&\Bigg\| \sum_{i=1}^{K_{[p_b^\ell]}} \pi_i^{[p_b^\ell]} k(s_{[p_a^\ell]i}) \{ f^{p^\ell}(s_{[p_I^\ell]i}, \{\tilde{n}^{*p}(s_{[p]i})\}_{p\in par(p^\ell)}) \}_{i=1}^{K_{[p_n^\ell]}}] \\
&- \int k^{p^\ell}(s_{[p_a^\ell]}) f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) ds_{[p_b^\ell]} \Bigg\|_\infty \\
\leq & C \Bigg\| \sum_{i=1}^{K_{[p_b^\ell]}} M_{[p_b^\ell]} [\{k^{p^\ell}(s_{[p_a^\ell]i}) f^{p^\ell}(s_{[p_I^\ell]i}, \{\tilde{n}^{*p}(s_{[p]i})\}_{p\in par(p^\ell)})\}_{i=1}^{K_{[p_b^\ell]}}] \phi_i \\
&- k^{p^\ell}(s_{[p_a^\ell]}) f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \Bigg\|_\infty \\
=& C \Bigg\| \sum_{i=1}^{K_{[p_b^\ell]}} M_{[p_b^\ell]} [\{k^{p^\ell}(s_{[p_a^\ell]i}) f^{p^\ell}(s_{[p_I^\ell]i}, \{\tilde{n}^{*p}(s_{[p]i})\}_{p\in par(p^\ell)})\}_{i=1}^{K_{[p_b^\ell]}}] \phi_i \\
&- k^{p^\ell}(s_{[p_a^\ell]}) f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]i})\}_{p\in par(p^\ell)}) \Bigg\|_\infty \\
&+ C \Big\| k^{p^\ell}(s_{[p_a^\ell]}) f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]i})\}_{p\in par(p^\ell)}) - \\
&k^{p^\ell}(s_{[p_a^\ell]}) f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \Big\|_\infty \\
=& C\zeta^{p_b^\ell}(K_{[p_b^\ell]}) + C \sum_{p\in par(p^\ell)} \left\| \tilde{n}^{*p}(s_{[p]}) - n^{*p}(s_{[p]}) \right\|_\infty^{\alpha^{p^\ell(p)}}
\end{aligned}
$$

where the first line follows from the definition of $\tilde{n}^{*p^\ell}$ and the second from the fact that the quadrature is exact over the interpolation method and that the integral is bounded by the sup norm of the integrand by the boundedness of the support $\mathcal{S}_{[p_b^\ell]}$. Line 3 follows from the triangle inequality. The first term in the final line uses that $k^{p^\ell}(s_{[p_a^\ell]i}) f^{p^\ell}(s_{[p_I^\ell]i}, \{\tilde{n}^{*p}(s_{[p]i})\}_{p\in par(p^\ell)}) \in \hat{\mathcal{G}}^{[p_b^\ell]}$ by the inductive hypothesis and Conditions 5(iv) and (v), while the second term comes from Conditions 5(i) and the fact that $k^{p^\ell}(s_{[p_a^\ell]})$ is bounded. The conclusion follows from the inequality that $(a+b)^\alpha \leq C_\alpha(a^\alpha + b^\alpha)$ for some $C_\alpha > 0$, for all $a,\ b \geq 0$. Finally, $\tilde{n}^{*p^\ell} \in \hat{\mathcal{G}}^{[p^\ell]}$ by Conditions 5(iv) (a) and (b). Thus, the inductive hypotheses hold for all p^ℓ, completing the proof of inequality (27) for all nodes.

To show inequality (6), first note that by the construction of function $\tilde{n}^{*p^\ell}$ and vector $\overrightarrow{n}^{*p^\ell}$, $\forall \ell$, $\forall p^\ell = 1 \dots P^\ell$, $\tilde{n}^{*p^\ell}(s_{[p^\ell]i}) = (\overrightarrow{n}^{*p^\ell})_i$ for all $i = 1 \dots K_{[p^\ell]}$, i.e., that $\tilde{n}^{p^\ell}$ is identical to the output of the algorithm at all points of evaluation.

As a result, I can construct an error bound as

$$\begin{aligned}
\left\| \sum_{k=1}^{K_{[p^\ell]}} (M_{[p^\ell]} \overrightarrow{n}^{*p^\ell})_k \phi_k - n^{*p^\ell} \right\|_\infty &= \left\| \sum_{k=1}^{K_{[p^\ell]}} (M_{[p^\ell]} \{\tilde{n}^{*p^\ell}(s_{[p^\ell]i})\}_{i=1}^{K_{[p^\ell]}})_k \phi_k - n^{*p^\ell} \right\|_\infty \\
&\leq \left\| \sum_{k=1}^{K_{[p^\ell]}} (M_{[p^\ell]} \{\tilde{n}^{*p^\ell}(s_{[p^\ell]i})\}_{i=1}^{K_{[p^\ell]}})_k \phi_k - \tilde{n}^{*p^\ell} \right\|_\infty \\
&\quad + \left\| \tilde{n}^{*p^\ell} - n^{*p^\ell} \right\|_\infty \\
&\leq C\zeta^{[p^\ell]}(K_{[p^\ell]}) + C(\zeta^{p^\ell}(K_{p^\ell})1_{p^\ell} + \\
&\quad \sum_{j \in dp(p^\ell)} \sum_{p \in j} (\zeta^{p(e)}(K_{p(e)})1_p)^{\alpha^{p,j}})
\end{aligned}$$

where the first term in the final final line follows from the fact that $\tilde{n}^{*p^\ell} \in \hat{\mathcal{G}}^{[p^\ell]}$ and 5(vi). □

Proof. of Lemma (9)

For each ℓ, for each $j = 1 \dots 2d_2 + o^\ell$, the proof of inequality (29) can be performed inductively along the directed graph starting at input nodes $p^\ell = 1 \dots 2d_2 + o^\ell$ for which $\left\| \tilde{n}_j^{p^\ell} - n_j^{p^\ell} \right\|_\infty = \left\| \tilde{n}_j^{p^\ell m} - n_j^{p^\ell m} \right\|_\infty = \left\| 1\{p^\ell = j\} - 1\{p^\ell = j\} \right\|_\infty = 0$ as $M_j^{p^\ell} = \{1\}$ for any node $p^\ell \in \{1 \dots 2d_2 + o^\ell\}$, and $\tilde{n}_j^{p^\ell m} = 1\{p^\ell = j\} \in \hat{\mathcal{G}}_d^{[p_j^{\ell m}];[p_{ej}^{\ell m}]}$ for all $m \in M_j^{p^\ell}$ by assumption (6)(iii), so the base case holds. For each p^ℓ, assume that inequality (29) holds for all $p \in dp(j, p^\ell)/p^\ell$, and that $\forall q \in q_j(p^\ell)$, $\forall m' \in M_j^q$, $\tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \in \hat{\mathcal{G}}_d^{[q_j^{m'}];[q_{ej}^{m'}]}$. Then, in the case where node p^ℓ has identity for a linear component, obtain bound

$$
\begin{aligned}
&\left\| \tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')} \right\|_\infty \\
&= \Bigg\| \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \cdot \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \\
&\quad - \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \cdot n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \Bigg\|_\infty \\
&\leq \left\| \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) - \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \right\|_\infty \cdot \\
&\quad \left\| \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right\|_\infty + \\
&\quad \left\| \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \right\|_\infty \left\| \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) - n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right\|_\infty \\
&\leq C\Big(\sum_{p\in par(p^\ell)} \left\| \tilde{n}^{*p}(s_{[p]}) - n^{*p}(s_{[p]}) \right\|_\infty^{\alpha_d^{p^\ell(p)}} + \sum_{(p,\tilde{m})\in dp(j,q,m')} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{s\in par(p)} (\upsilon^p)^{\alpha_d^{p(s)}}) \Big) \\
&\leq (\zeta_d^{p_{dj}^\ell}(K_{[p_{dj}^\ell]}) \cdot 0 + C\Big(\sum_{p\in par(p^\ell)} (\upsilon^p)^{\alpha_d^{p^\ell(p)}} + \sum_{(p,\tilde{m})\in dp(j,q,m')} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{s\in par(p)} (\upsilon^p)^{\alpha_d^{p(s)}}) \Big) \\
&= C \sum_{(p,\tilde{m})\in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q\in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}})
\end{aligned}
$$

Here, the second inequality holds by the inductive hypothesis and condition (6)(ii) and (iii) to ensure that exact and approximate derivatives lie in boundedness of classes for all p^ℓ, and condition (6)(i), and the third by Lemma (7).

In the case where node p^ℓ has an integral operator for a linear component,

obtain bound $\left\|\tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')}\right\|_\infty$

$$
\begin{aligned}
=&\Bigg\| \sum_{i=1}^{K_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}} \pi_i^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} k^{p^\ell}(s_{[p_a^\ell],i}) \cdot \\
&\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell],i}, \{\tilde{n}^{*p}(s_{[p],i})\}_{p\in par(p^\ell)}) \cdot \tilde{n}_j^{qm'}(s_{[q_j^{m'}]i}, s'_{[q_{ej}^{m'}]}) \\
&- \int k^{p^\ell}(s_{[p_a^\ell]}) \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \cdot n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) ds_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} \Bigg\|_\infty \\
\leq& \zeta_{dK}^{p^\ell jm} + \Bigg\| \int k^{p^\ell}(s_{[p_a^\ell]}) \Bigg[\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \cdot \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \\
&- \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \cdot n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \Bigg] ds_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} \Bigg\|_\infty \\
\leq& \zeta_{dK}^{p^\ell jm} + C \Bigg\| k^{p^\ell}(s_{[p_a^\ell]}) \Bigg[\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \\
&- \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \Bigg] \Bigg\|_\infty \left\| \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right\|_\infty \\
&+ C \left\| k^{p^\ell}(s_{[p_a^\ell]}) \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \right\|_\infty \cdot \\
&\left\| \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) - n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right\|_\infty \\
\leq& \zeta_{dK}^{p^\ell jm} + C(\sum_{p\in par(p^\ell)} \left\| \tilde{n}^{*p}(s_{[p]}) - n^{*p}(s_{[p]}) \right\|_\infty^{\alpha_d^{p(q)}} + \\
&\sum_{(p,\tilde{m})\in dp(j,q,m')} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{s\in par(p)} (\upsilon^p)^{\alpha_d^{p(s)}})) \\
\leq& (\zeta_{dK}^{p^\ell jm} \cdot 1 + C(\sum_{p\in par(p^\ell)} (\upsilon^p)^{\alpha_d^{p^\ell(p)}} + \sum_{(p,\tilde{m})\in dp(j,q,m')} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{s\in par(p)} (\upsilon^p)^{\alpha_d^{p(s)}})) \\
=& C \sum_{(p,\tilde{m})\in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q\in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}})
\end{aligned}
$$

where the first inequality follows from assumption (6)(iii) and (iv), the second by boundedness of support of $\mathcal{S}_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}$, the third by the inductive hypothesis and assumptions (6)(i), (ii), and (iii), and the fourth by Lemma (7).

Applying this inequality to the definition of $\tilde{n}_j^{p^\ell}$, obtain $\left\|\tilde{n}_j^{p^\ell} - n_j^{p^\ell}\right\|_\infty$

$$
\begin{aligned}
&= \left\| \sum_{m \in M_j^{p^\ell}} \tilde{n}_j^{p^\ell m(q,m')} - \sum_{m \in M_j^{p^\ell}} n_j^{p^\ell m(q,m')} \right\|_\infty \\
&\leq \sum_{m \in M_j^{p^\ell}} \left\| \tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')} \right\|_\infty \\
&\leq C \sum_{m \in M_j^{p^\ell}} \sum_{(p,\tilde{m}) \in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}})
\end{aligned}
$$

Next, by assumptions (6)(iii) and (iv), $\tilde{n}_j^{p^\ell m(q,m')} \in \hat{\mathcal{G}}_d^{[p_j^{\ell m}];[p_{ej}^{\ell m}]}$ in either of the above cases, so the inductive hypothesis holds and the claim is true for all p^ℓ.

To show inequality (30), note that $\forall \ell$, $\forall j = 1 \dots 2d_2$, by condition (3) ensuring that the argument of an output node P^ℓ is a function only of $\mathcal{S}_{[\ell^o]}$ and the construction of sets $[P_{ej}^{\ell m}]$ as subsets of $[j]$, function $\tilde{n}_j^{P^\ell} = \sum_{m \in M_j^{p^\ell}} \tilde{n}_j^{P^\ell m}(s_{[P_j^{\ell m}]}, s'_{[P_{ej}^{\ell m}]})$ is defined over $\mathcal{S}_{[\ell^o]} \times \mathcal{S}_{[j]}$. Following Algorithm (1) and the application of the chain rule, it can therefore be seen that $[\mathcal{F}^\ell_{\overrightarrow{g}_j(in)}(\Pi^{[j]})^{-1}]_{ik} = \tilde{n}_j^{P^\ell}(s_{[\ell^o],k}, s'_{[j],i})\ \forall i = 1 \dots K_{[j]}, k = 1 \dots K_{[\ell^o]}$. As a result, I have that

$$
\begin{aligned}
&\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} (M_{[\ell^o]} \mathcal{F}^\ell_{\overrightarrow{g}_j(in)} (\Pi^{[j]})^{-1} M^*_{[j]})_{ik} \phi_i \phi_k - n_j^{P^\ell} \right\|_\infty = \\
&\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} ((M_{[\ell^o]} \otimes M_{[j]}) \{ \{ \tilde{n}_j^{P^\ell}(s_{[\ell^o],r_1}, s'_{[j],r_2}) \}_{r_2=1}^{K_{[j]}} \}_{r_1=1}^{K_{[\ell^o]}})_{ik} \phi_i \phi_k - n_j^{P^\ell} \right\|_\infty \leq \\
&\left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} ((M_{[\ell^o]} \otimes M_{[j]}) \{ \{ \tilde{n}_j^{P^\ell}(s_{[\ell^o],r_1}, s'_{[j],r_2}) \}_{r_2=1}^{K_{[j]}} \}_{r_1=1}^{K_{[\ell^o]}})_{ik} \phi_i \phi_k - \tilde{n}_j^{P^\ell} \right\|_\infty + \left\| \tilde{n}_j^{P^\ell} - n_j^{P^\ell} \right\|_\infty \\
&\leq \sum_{m \in M_j^{P^\ell}} \left\| \sum_{k=1}^{K_{[\ell^o]}} \sum_{i=1}^{K_{[j]}} ((M_{[\ell^o]} \otimes M_{[j]}) \{ \{ \tilde{n}_j^{P^\ell m}(s_{[P_j^{\ell m}],r_1}, s'_{[p_{ej}^{\ell m}],r_2}) \}_{r_2=1}^{K_{[j]}} \}_{r_1=1}^{K_{[\ell^o]}})_{ik} \phi_i \phi_k - \tilde{n}_j^{P^\ell m} \right\|_\infty \\
&\qquad + \left\| \tilde{n}_j^{P^\ell} - n_j^{P^\ell} \right\|_\infty \\
&\leq C \sum_{m \in M_j^{P^\ell}} (\zeta^{\ell j m}(K_{[\chi+1] \times [j]}) + \sum_{(p,\tilde{m}) \in dp(j,P^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}}))
\end{aligned}
$$

where the first line follows from the construction of $\tilde{n}_j^{P^\ell}$, the second and third by the triangle inequality, and the fourth by assumption (6) (v) and inequality (29) applied to the derivative of node P^ℓ with respect to input j.

To show inequality (31), note that by condition (4)(b), $\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^{\ell}(out))}}$ is a diagonal matrix with entries given by $\tilde{n}^{P^{\ell}}_{j(out)}(s_{[\ell^o],k})\ \forall k = 1 \dots K_{[\ell^o]}$, which are bounded away from 0 by Condition (6)(vi) and so applying the same logic as for condition (30), $\left[(\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^{\ell}(out))}})^{-1}\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}}(\Pi^{[j]})^{-1}\right]_{ik} = \tilde{n}^{P^{\ell}}_{j}(s_{[\ell^o],k}, s'_{[j],i})/\tilde{n}^{P^{\ell}}_{j(out)}(s_{[\ell^o],k})$ $\forall i = 1 \dots K_{[j]}, k = 1 \dots K_{[\ell^o]}$, and so

$$\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}(M_{[\ell^o]}(\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^{\ell}(out))}})^{-1}\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}}(\Pi^{[j]})^{-1}M^{*}_{[j]})_{ik}\phi_i\phi_k - n^{P^{\ell}}_{j}/n^{P^{\ell}}_{j(out)}\right\|_{\infty} =$$

$$\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}((M_{[\ell^o]}\otimes M_{[j]})\{\{\tilde{n}^{P^{\ell}}_{j}(s_{[\ell^o],r_1}, s'_{[j],r_2})/\tilde{n}^{P^{\ell}}_{j(out)}(s_{[\ell^o],r_1})\}_{r_2=1}^{K_{[j]}}\}_{r_1=1}^{K_{[\ell^o]}})_{ik}\phi_i\phi_k - n^{P^{\ell}}_{j}/n^{P^{\ell}}_{j(out)}\right\|_{\infty} \leq$$

$$\sum_{m\in M^{P^{\ell}}_{j}}\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}((M_{[\ell^o]}\otimes M_{[j]})\{\{\tilde{n}^{P^{\ell}m}_{j}(s_{[\ell^o],r_1}, s'_{[j],r_2})/\tilde{n}^{P^{\ell}}_{j(out)}(s_{[\ell^o],r_1})\}_{r_2=1}^{K_{[j]}}\}_{r_1=1}^{K_{[\ell^o]}})_{ik}\phi_i\phi_k - \tilde{n}^{P^{\ell}m}_{j}/\tilde{n}^{P^{\ell}}_{j(out)}\right\|_{\infty} +$$

$$\left\|\tilde{n}^{P^{\ell}}_{j}/\tilde{n}^{P^{\ell}}_{j(out)} - n^{P^{\ell}}_{j}/n^{P^{\ell}}_{j(out)}\right\|_{\infty} \leq$$

$$C\sum_{m\in M^{P^{\ell}}_{j}}(\zeta^{\ell jm}_{d}(K_{[\chi+1]\times[j]}) + \left\|\tilde{n}^{P^{\ell}}_{j} - n^{P^{\ell}}_{j}\right\|_{\infty} + \left\|\tilde{n}^{P^{\ell}}_{j(out)} - n^{P^{\ell}}_{j(out)}\right\|_{\infty}) \leq$$

$$C(\sum_{m\in M^{P^{\ell}}_{j}}\zeta^{\ell jm}_{d}(K_{[\chi+1]\times[j]}) + \sum_{m\in M^{P^{\ell}}_{j}}\sum_{(p,\tilde{m})\in dp(j,P^{\ell},m)}(\zeta^{pj\tilde{m}}_{dK}1_p + \sum_{q\in par(p)}(\upsilon^{p})^{\alpha^{p(q)}_{d}})+$$

$$\sum_{m\in M^{P^{\ell}}_{j(out)}}\sum_{(p,\tilde{m})\in dp(j(out),P^{\ell},m)}(\zeta^{pj(out)\tilde{m}}_{dK}1_p + \sum_{q\in par(p)}(\upsilon^{p})^{\alpha^{p(q)}_{d}}))$$

where the first equality follows from the above discussion, the first inequality by the triangle inequality, the second inequality by assumption (6) (vi) and using that $\tilde{n}^{P^{\ell}}_{j(out)}$ and $n^{P^{\ell}}_{j(out)}$ are bounded below, and the last by inequality (29) applied to the derivative of node P^{ℓ} with respect to inputs j and $j(out)$ respectively. □

Proof. of Theorem (10)

For each equilibrium condition ℓ, the functional derivative of $\mathcal{F}^{\ell}$ with respect to g_j for $j = 1 \dots 2d_2$ is the sum of 2 components, a derivative with respect to g_j as an input variable and as an output variable. Under Conditions (3) and (4) (b)(ii), the functional derivative with respect to g_j as an input, corresponding to nodes $p^{\ell} = j \in 1 \dots 2d_2$ takes the form of a compact integral operator with kernel function $n^{P^{\ell}_{j}}$ (where compactness holds because sup norm approximability of the kernel function by a countable basis, which holds by Conditions (5) and (6), implies operator norm approximability by a finite matrix by Young's inequality, which implies compactness). The functional derivative with respect to g_j as an output, corresponding to the nodes $p^{\ell} \in [2d_2 + 1, 2d_2 + o^{\ell}]$ such that $j = j(p^{\ell})$,

takes the form of a scalar multiple of the identity. Since the functional derivative is a sum of these two components, each row ℓ of this form is consistent with asymptotic diagonality. For rows ℓ satisfying condition (4) (b)(i), the same holds subsequent to division by $n_{2d_2+1}^{P^\ell}$, a fixed nonlinear function.

Given this form, Algorithm (1) constructs an approximation $(\tilde{B}^K, \tilde{A}^K)$ of functional derivatives (B, A) of equilibrium conditions $\mathcal{F}$ such that the integral operator components on each block (ℓ, j) have kernel which is approximated in sup norm using a set of basis functions on $\mathcal{H}^K$ at rate $\epsilon_K^{(\ell,j)}$, and so converges in operator norm by Young's inequality and the boundedness of domain also at rate $\epsilon_K^{(\ell,j)}$, and the identity components on this block are approximated by an identity matrix on $\mathcal{H}^K$, giving a representation $(\tilde{B}^K, \tilde{A}^K) = (\tilde{B_C}^K, \tilde{A}_C^K) + (B_I^K, A_I^K)$ in terms of integral and identity components. By Lemma (9) and by bounding the operator norm of the approximation errors in $(\left\|\tilde{B}_C^K - B_C^K\right\|_{op}, \left\|\tilde{A}_C^K - A_C^K\right\|_{op})$ by the sum of the operator norm of the errors on each block $\max(\left\|\tilde{B}_C^K - B_C^K\right\|_{op}, \left\|\tilde{A}_C^K - A_C^K\right\|_{op}) \leq \sum_{\ell,j} \epsilon_K^{(\ell,j)} \leq d_2^2 \epsilon_K$. As a result, by Theorem 1 of Childers [2018],

$$\sup_{\|f\|_{\mathcal{H}_x}=1} \left\|(\tilde{h}_x^K + h_x^{K\perp} - h_x)f\right\| \leq O(d_2^2 \epsilon_K)$$

$$\sup_{\|f\|_{\mathcal{H}_x}=1} \left\|(\tilde{g}_x^K + g_x^{K\perp} - g_x)f\right\| \leq O(d_2^2 \epsilon_K)$$

By orthogonality, for any $f \in \mathcal{H}^K$ $\tilde{h}_x^K f = (\tilde{h}_x^K + h_x^{K\perp})f$ and $\tilde{g}_x^K f = (\tilde{g}_x^K + g_x^{K\perp})f$, and so the claim of the theorem follows. □

Proof. of Theorem (13). First, it is demonstrated that the proposed modification of Algorithm (1) produces numerically equivalent results to application of the algorithm to a model with g and Lg solved out, then the conditions of Theorem (10) are verified for this model.

Applying the implicit function theorem to first order conditions equation (7) (represented as $F^{(j)}$), with derivatives expressed in Section (B.1), obtain Taylor expansion

$$g_t(X, \epsilon) \approx g^*(X, \epsilon) + (\frac{\partial}{dg_t(X, \epsilon)} F^{(j)})^{-1}$$

$$\left(\frac{\partial}{dP_t} F^{(j)}[P_t - P^*] + \frac{\partial}{dV_{t+1}} F^{(j)}[V_{t+1}(.) - V^*] + \frac{\partial}{dP_{t+1}} F^{(j)}[P_{t+1} - P^*]\right)$$

By Condition (12)(i), $(\frac{\partial}{dg_t(X,\epsilon)} F^{(2)})^{-1}$ exists and is a bounded multiplication operator, and and each summand is a compact operator (either an integral operator or over finite dimensional inputs). Consider the model defined by replacing $g_t(X, \epsilon)$ with this Taylor expansion wherever $g_t(X, \epsilon)$ appears and removing equation (7). By the chain rule, for any differentiable F,

the derivative of $F(g(V,P,P'))$ with respect to (V,P,P') at the steady state is $\frac{\partial}{dg}F(g(V^*,P^*,P^*))\frac{\partial g}{d(V,P,P')}[(V,P,P')-(V^*,P^*,P^*)]$, the same as the derivative of $F(g(V^*,P^*,P^*)+\frac{\partial g}{d(V,P,P')}[(V,P,P')-(V^*,P^*,P^*)])$, so this substitution produces derivatives identical to those produced by solving for g explicitly.

By the choice of $I_{K_\epsilon \times K_x}$ in place of $M^{\mathcal{X}}_{K_x} \otimes M^{\epsilon}_{K_\epsilon}$ in the numerical derivative of equation (7) and in columns j corresponding to $g(X,\epsilon)$ or $Lg(X,\epsilon)$ in step (5) of the algorithm, and treating g_t as an input variable in equation (6), the linearized model produced by applying Algorithm (1) to the model with $g_t(.)$ substituted out is numerically identical to that produced by applying the proposed modified version of Algorithm (1), then solving linearized equation (7) for g_t to obtain vector $\frac{dg(X,\epsilon)}{d(P,V',P')} := (\frac{\partial}{dg_t(X,\epsilon)}\tilde{F}^{(j)})^{-1}\left(\begin{array}{ccc} \frac{\partial}{dP_t}\tilde{F}^{(j)}\Pi^{-1}_{[P]}M^*_{[P]} & \frac{\partial}{dV_{t+1}}\tilde{F}^{(j)}\Pi^{-1}_{[V]}M^*_{[V]} & \frac{\partial}{dP_{t+1}}\tilde{F}^{(j)}\Pi^{-1}_{[P]}M^*_{[P]} \end{array} \right)$ and multiplying the derivative with respect to $g(X,\epsilon)$ in all equations by this solution and adding to the corresponding blocks. One may then apply the identity $Lg' = g$ to solve out Lg from all equations.

Due to the above equivalence and applying numerical equivalence result from lemma (28), application of the linear rational expectations solution algorithm will produce matrices h^K_x, g^K_x with subcomponents $h^K_{x(-j)}$, $g^K_{x(-j)}$ corresponding to deleting the rows and columns corresponding to Lg and g identical to those of the solved out model. The row of g^K_x corresponding to output $g(X,\epsilon)$ and the row of h^K_x corresponding to output Lg' are both then given by vector postmultiplying $\frac{dg(X,\epsilon)}{d(P,V',P')}$ by the subcomponents of $g^K_{x(-j)}$, $h^K_{x(-j)}$, and $g^K_{x(-j)}h^K_{x(-j)}$ producing outputs P_{1t}, P_{t+1}, and V_{t+1}. The columns of h^K_x and g^K_x corresponding to input Lg are 0.

Supposing that the solved out model satisfies the conditions of Theorem (10) $\tilde{h}^K_{x(-j)}$, $\tilde{g}^K_{x(-j)}$ converge in operator norm, at rate $O(\max\{\zeta_{K^2_\epsilon \times K^2_x}, \zeta_K, \zeta_P\})$.

By equivalent calculations to those for $\tilde{h}^K_{x(-j)}$, $\tilde{g}^K_{x(-j)}$, $(M^{\mathcal{X}}_{K_x} \otimes M^{\epsilon}_{K_\epsilon})\frac{dg(X,\epsilon)}{d(P,V',P')}$ converges in operator norm at rate $O(\max\{\zeta_{K^2_\epsilon \times K^2_x}, \zeta_K, \zeta_P\})$, and so the composition with $\tilde{h}^K_{x(-j)}$, $\tilde{g}^K_{x(-j)}$ does as well. Therefore, the modified step (6) produces operators with the claimed convergence rates.

I now verify the conditions of Theorem (10) for $\tilde{h}^K_{x(-j)}$, $\tilde{g}^K_{x(-j)}$.

First, the model with transformations takes the structure defined in Condition (3), with all maps taking the structure of integral operators and nonlinear transforms using state variables which are functions on subsets of $\mathbb{R}^d$, with $\mathcal{X}$, with by the structure of the equations and Condition (12).

Condition (4)(i) is satisfied as all constituent functions are bounded, measurable (by continuity), differentiable, and Hölder continuous in their arguments (with $\alpha^p \geq 1$) by Condition (12).

Condition (4)(ii) is satisfied for all equations after the substitution out of $g(X,\epsilon)$, with part (i) satisfied by all equations except the lag definition equations (13), which satisfy part (ii) instead.

Condition (5) (i)-(iii) are satisfied with all function classes $\mathcal{G}$ chosen to be 1-Hölder functions due to the Hölder conditions of Condition (12) and the fact that this property is preserved by composition, addition, multiplication, division

when bounded away from 0, and marginalization to sub-arguments. Parts (iv) and (v) hold by Condition (12)(ii) on the interpolation and the choice of 1-Hölder functions as the class $\hat{\mathcal{G}}$. Part (vi) holds by Condition (12)(ii) and (iii), with rates given by the steady state assumption and rates of tensor product interpolation schemes over each corresponding function class, the slowest of which will be of the order of $\zeta_{K_\epsilon^2 \times K_x^2}$ or ζ_P.

Condition (6) holds with all function classes $\mathcal{G}_d$ and $\hat{\mathcal{G}}_d$ chosen to be 1-Hölder functions due to the Hölder conditions on derivatives of Condition (12) the composition properties of Hölder functions, and the properties of the interpolation scheme, noting that for bounded functions the rate of interpolation error $\zeta_{dk}^{p^\ell jm}$ is bounded by the order of the rate of sup norm error. Rates of convergence are in all cases those for 1-Holder functions of the corresponding dimensions. Part (vi) holds vacuously.

As a result of the above the conditions of lemmas (7) and (9) hold, with rates on the order of $O(\max\{\zeta_{K_\epsilon^2 \times K_x^2}, \zeta_K, \zeta_P\})$ for each equation, and so the system of functional derivatives can be assured to converge in operator norm at this rate. The above arguments demonstrate that the conditions of lemmas (7) and (9) apply also to $(M_{K_x}^{\mathcal{X}} \otimes M_{K_\epsilon}^{\epsilon})^{\frac{dg(X,\epsilon)}{d(P,V',P')}}$, so this also converges in operator norm at the same or better rates.

The conditions of Theorem (10) hold by the operator norm convergence of the derivatives and by the assumption that the existence and continuity conditions of Childers [2018] hold, so the approximate linearized solution using the transformed variables also converges at the rate $O(\max\{\zeta_{K_\epsilon^2 \times K_x^2}, \zeta_K, \zeta_P\})$. Together with the numerical equivalence to the modified model, this proves the result. □

Proof. of Lemma (14). The first part of this proceeds by first verifying the conditions of Theorem (10) for the proposed model using a transformed histogram representation via the results of Lemma (24), then applying Corollary (26) to derive bounds for the representation in terms of the evenly spaced histogram representation.

First, condition (3) representation can be seen by inspection to hold for the system of equations defining the model. Condition (4)(a) can be seen to hold by noting that all functions defining each equilibrium condition are Hölder of order at least 1 (noting that the pointwise max function is 1-Lipschitz, and the CRRA function and its derivative are uniformly infinitely differentiable over any neighborhood not containing zero). Condition (4)(b)(ii) is seen to hold for all equations except for Euler equation, for which condition (4)(b)(i) holds.

For the choice of function spaces and interpolation scheme, while evenly spaced histograms would be sufficient to show the convergence of the steady state in sup norm, it will be necessary for the next step to demonstrate convergence with respect to a particular choice of transformed histogram representation. The derivative of map $c(w, \ell(w), R)$ with respect to $\ell(.)$ at $\ell^*(w), R^*$ is $1\{\ell^*(w)^{-\frac{1}{\gamma}} > w - \frac{\mathrm{a}}{R^*}\} \cdot [.]$. Because $\ell^*(w)^{-1/\gamma}$ can be shown to be a strictly monotone function, there exists a unique value w^* such that $\ell^*(w^*)^{-\frac{1}{\gamma}} = w^* - \frac{\mathrm{a}}{R^*}$. By

the assumed uniform convergence of $\tilde{\ell}^*(w)$ and $\tilde{R}^*$ at rate $O(K^{-1})$ and either the monotonicity or an $o(K^{-1})$ convergence rate, for large enough K there exists a $\tilde{w}^*$ such that $\ell^*(w_i)^{-\frac{1}{\gamma}} > w_i - \frac{\underline{a}}{\tilde{R}^*}$ for all $w_i > \tilde{w}^*$ and the reverse holds for all $w_i < \tilde{w}^*$, and $\tilde{w}^* - w^* = O(K^{-1})$. There therefore exists a transformation $\tau^K(w)$ satisfying condition (22), constructed by stretching the areas above and below $\tilde{w}^*$ to reach w^*, and so the corresponding transformed histogram representation of the steady state satisfies $1\{\ell^*(w)^{-\frac{1}{\gamma}} > w - \frac{\underline{a}}{\tilde{R}^*}\} = 1\{\tilde{\ell}^*(w)^{-\frac{1}{\gamma}} > w - \frac{\underline{a}}{\tilde{R}^*}\}$. Let this be the approximation scheme used in verifying conditions (5) and (6) .

Condition (5)(i) may be seen to hold by the Hölder continuity of all nodes. Define $\mathcal{PR}_{1,1}^{\alpha,N}(\mathcal{X})$ as the class of functions $f(x) : \mathcal{X} \subset \mathbb{R}^d \to \mathbb{R}$ with weak partial derivatives $\frac{\partial}{dx_i} f(x)\ \forall i = 1 \dots d$ uniformly piecewise regular with exponent α, in the sense of definition (20), and for all argument sets $[p]$ define the class $\mathcal{G}^{[p]} := \mathcal{PR}_{1,1}^{1,N}(\mathcal{S}_{[p]})$ of functions with derivatives which are piecewise regular with Hölder exponent 1 in all arguments. By the corresponding properties for Hölder continuous functions, the class $\mathcal{PR}_{1,1}^{1,N}(\mathcal{X})$ can be seen to be stable under addition, composition, multiplication, and, if the numerator is bounded away from 0, division, and so condition (5)(ii) on stability of class membership also holds for classes $\mathcal{G}^{[p]}$. Condition (5)(iii) that the steady state functions are members of this class may be verified using known properties of the steady state of consumption savings models with borrowing constraints and incomplete markets; see, e.g. Ljungqvist and Sargent [2012, Ch 17], Carroll and Kimball [1996], Carroll [2004]. In particular, these imply that the steady state wealth distribution has bounded support, and that the consumption function is concave, and so, on this support, bounded, and so by the law of motion for cash on hand and the bounded support of the income shock, the distribution of cash on hand has finite support. The consumption function can be shown to be piecewise twice differentiable and so a member of $\mathcal{PR}_{1,1}^{1,N}$ on its support. Because the distribution g is smooth, the Kolmogorov forward equation maps any distribution $m_t(.)$ to a distribution with smooth density $m_{t+1}(.)$ and so the steady state must also be smooth. $Lm^* = m^*$ and $L\ell^* = \ell^*$ in steady state, and so all input functions satisfy the desired regularity conditions. Using the transformed histogram interpolation, condition (5)(iv) holds by the preservation of classes $\mathcal{G}^{[p]}$ under summation and marginalization. Condition (5)(v)(i) and (ii) ask for rate of convergence for interpolation and quadrature over these function classes: by Lemma (24) and the fact that $\mathcal{PR}_{1,1}^{1,N}$ contains $\mathcal{PR}_{0,0}^{1,N}$, the sup norm convergence rate for transformed histogram interpolation of a function in dimension d must be $\zeta_K = O(K^{-\frac{1}{2d}})$. Condition (5)(v) holds by the assumption on the convergence rate of the steady state with rate $O(K^{-1})$.

As a result, the conditions of Lemma (7) hold and steady state nodes all converge at rate $O(K^{-\frac{1}{2}})$.

Condition (6) holds due to a set of considerations similar to those above for condition (5), except this time applied to the derivatives rather than the values of the functions defining the equilibrium conditions. Condition (6)(i) requires some special consideration; for all nodes except that corresponding to the mappings $c(w, \ell(.), R)$, this bound holds by Hölder continuity. For the nodes $c(w, \ell(.), R)$,

the derivatives with respect to ℓ and R are $\frac{-1}{\gamma}1\{\ell^*(w)^{-\frac{1}{\gamma}} > w - \frac{\underline{a}}{R^*}\}\ell^*(w)^{-\frac{1}{\gamma}-1}[.]$ and $-1\{\ell^*(w)^{-\frac{1}{\gamma}} > w - \frac{\underline{a}}{R^*}\}\frac{\underline{a}}{R^{*2}}[.]$ respectively, which are not Hölder with respect to arbitrary changes in ℓ^*,R^*. However, in the proof of Lemma (9), this condition is only used to bound the sup norm difference between these derivatives and the versions using the approximate steady state values $\frac{-1}{\gamma}1\{\tilde{\ell}^*(w)^{-\frac{1}{\gamma}} > w - \frac{\underline{a}}{\tilde{R}^*}\}\tilde{\ell}^*(w)^{-\frac{1}{\gamma}-1}[.]$ and $-1\{\tilde{\ell}^*(w)^{-\frac{1}{\gamma}} > w - \frac{\underline{a}}{\tilde{R}^*}\}\frac{\underline{a}}{\tilde{R}^{*2}}[.]$, which by construction of the transformed histogram representation does converge in sup norm because the indicator function is calculated exactly. Conditions (6) (ii) through (vi) hold using as function classes piecewise regular functions with piecewise Hölder exponent at least 1 for all arguments and transformed histogram representations for all interpolation and quadrature, using the composition properties of these classes (noting that discontinuities are always located at fixed location w^*). For part (vi), note that the derivative of the Euler equation with respect to $\ell(.)$ is the only equation satisfying Condition (4)(b)(i). Boundedness away from 0 is satisfied by the fact that the steady state consumption function is always nonzero and monotone and so is the expected marginal utility of wealth. Dividing through by this derivative produces a derivative with respect to $\ell_{t+1}(w)$ which is a kernel integral operator in class $\mathcal{PR}^{1,4}_{0,0}$, due to the presence of discontinuities with respect to w and w', and so transformed histograms converge at rate $O(K^{-\frac{1}{4}})$ on this term. All other derivatives also lie in subsets of this class and so converge at least this quickly. As a result, by Lemma (9) with the specified modification holds and the functional derivatives of the model converge at rate $O(K^{-\frac{1}{4}})$, and by the assumptions on the properties of the model as a whole, the solution operators converge by the result of Theorem (10).

Finally, the claimed result holds for the representation of the solution operators constructed by Algorithm (1) and the corresponding impulse response functions, using regularly spaced histograms, by application of Corollary (26). □

C.2 Proofs of Claims from Appendices

Proof of Lemma (17)

Proof. Let $f(x_1, \ldots, x_d) \in \mathcal{F}$, Letting i be a multi-index, $[\vec{f}]_i = \frac{1}{\sqrt{K}} f(s_{i_1}, s_{i_2,}, \ldots, s_{i_d})$ and $\varphi_i(x_1, \ldots, x_d) = \prod_{z=1\ldots d} \phi_{i_z}(x_z)$. $\hat{f} = \sum_{i_1=1}^{K_1} \ldots \sum_{i_d=1}^{K_d} [\vec{f}]_i \varphi_i(.)$ is the histogram approximation of f. The sup norm accuracy of this approximation is given by

$$\begin{aligned} \sup_{s\in[0,1)^d} \left|\hat{f}(s) - f(s)\right| &\leq \sup_{s\in[0,1)^d} \left|\hat{f}(s) - \sum_{i_1=1}^{K_1} \ldots \sum_{i_d=1}^{K_d} \langle f, \varphi_i\rangle \varphi_i(.)\right| + \sup_{s\in[0,1)^d} \left|\sum_{i_1=1}^{K_1} \ldots \sum_{i_d=1}^{K_d} \langle f, \varphi_i\rangle \varphi_i(.) - f\right| \\ &= (I) + (II) \end{aligned}$$

where (I) is the quadrature error and (II) is the projection error.

In one dimension, the bound on (II) is given explicitly in Nickl (2013) Propositions 7.iii and 9.iii, and is in this case $O(\frac{1}{K_z})$. Applying the method of proof

provided there in the multidimensional case, I have, for any $s \in [0,1)^d$ with index j corresponding to the (unique) basis function at which $\varphi_j(s) \neq 0$

$$\begin{aligned}
\left| \sum_{i_1=1}^{K_1} \cdots \sum_{i_d=1}^{K_d} \langle f, \varphi_i \rangle \varphi_i(s) - f(s) \right| &= |\langle f, \varphi_j \rangle \varphi_j(s) - f(s)| \\
&= \sqrt{K} \int \varphi_j(y)(f(y) - f(s))dy \\
&= \sqrt{K} \int \varphi_j(s+u)(f(s+u) - f(s))du \\
&\leq \sqrt{K} \int |\varphi_j(s+u)| \, |f(s+u) - f(s))| \, du \\
&\leq \sup_{u \in [\frac{-1}{2K_z}, \frac{1}{2K_z})^d} |f(s+u) - f(s))| \sqrt{K} \int |\varphi_j(s+u)| \, du \\
&= \sup_{u \in [\frac{-1}{2K_z}, \frac{1}{2K_z})^d} |f(s+u) - f(s))| \\
&\leq L(\frac{\sqrt{d}}{2K_z})^{\min\{\alpha,1\}} = O(d^{\frac{\min\{\alpha,1\}}{2}} K^{\frac{-\min\{\alpha,1\}}{d}})
\end{aligned}$$

where the first line follows from disjoint support, the second from the integral of φ_j, the third by a change of variables, the sixth again by the integral of φ_j, and the last by the Hölder condition, where L is the Lipschitz constant and I used that the maximum distance between two points in a hypercube of dimension d with sides of length c is $c\sqrt{d}$.

Using the finite support of the basis and the fact that the functions are piecewise constant, obtain

$$\begin{aligned}
(I) &= \sup_i \sup_{s \in [0,1)^d} \left| ([\vec{f}]_i - \langle f, \varphi_i \rangle) \varphi_i(.) \right| \\
&= \sup_i |f(s_{i_1}, s_{i_2},, \ldots, s_{i_d}) - \langle f, \varphi_i \rangle \varphi_i(.)|
\end{aligned}$$

This is exactly controlled by the first line of the bound for (II) shown above for the particular subset of points on the grid, and so is likewise bounded by $\frac{L\sqrt{d}}{2K_z} L(\frac{\sqrt{d}}{2K_z})^{\min\{\alpha,1\}}$.

Combining the bounds for (I) and (II) gives the result. □

Proof of Lemma (18)

Proof. By definition of the tensor product interpolation and the bound (4.26) in BG on the sup norm of each layer, for sufficiently large $n$so that the order of

the polynomials is constant on all neglected terms,

$$\begin{aligned}
\|u - u_n^{p,\infty}\|_\infty &\leq \sum_{|\ell|_\infty > n} \|u_\ell\|_\infty \leq 0.5585^d c(p)|u|_{p+1,\infty} \sum_{|\ell|_\infty > n} 2^{-|\ell\cdot(p+1)|_1} \\
&\leq 0.5585^d c(p)|u|_{p+1,\infty}(\sum_{\ell}(2^{p+1})^{-|\ell|_1} - \sum_{|\ell|_\infty \leq n}(2^{p+1})^{-|\ell|_1}) \\
&= 0.5585^d c(p)|u|_{p+1,\infty}((2^{p+1}-1)^{-d} - (\sum_{i=1}^{n}(2^{p+1})^{-i})^d) \\
&= 0.5585^d c(p)|u|_{p+1,\infty}(2^{p+1}-1)^{-d}(1-(1-(2^{p+1})^{-n})^d) \\
&\leq d0.5585^d c(p)|u|_{p+1,\infty}(2^{p+1}-1)^{-d}(2^{p+1})^{-n}
\end{aligned}$$

The error per function evaluation comes from the fact that $N = (2^n - 1)^d = O(2^{nd})$ function evaluations are needed to construct the tensor product interpolation, so $n = O(\frac{\log_2 K}{d})$. □

Proof of Lemma (19)

Proof. By the layerwise representation of u and bound (4.26) in BG, and the finiteness of the set of terms of order less than p

$$\begin{aligned}
\left\|u - u_n^{p,\{1,1\}}\right\|_\infty &\leq \sum_{\max\{|\ell^{d_1}|_1,|\ell^{d_2}|_1\} > n+\frac{d}{2}-1} \|u_\ell\|_\infty \qquad (34) \\
&\leq 0.5585^d c(p)|u|_{p+1,\infty} \sum_{\max\{|\ell^{d_1}|_1,|\ell^{d_2}|_1\} > n+\frac{d}{2}-1} 2^{-|\ell\cdot(p+1)|_1} + O(2^{-(p+1)n})
\end{aligned}$$

Using the fact that there are $\begin{pmatrix} i-1 \\ d-1 \end{pmatrix}$ ways to form the sum i with d non-negative integers, I can split up the final sum as

$$\begin{aligned}
\sum_{\max\{|\ell^{d_1}|_1,|\ell^{d_2}|_1\} > n+\frac{d}{2}-1} 2^{-|\ell\cdot(p+1)|_1} &= 2\sum_{i=n+\frac{d}{2}}^{\infty}\sum_{j=\frac{d}{2}}^{\infty} 2^{-(p+1)(i+j)} \begin{pmatrix} i-1 \\ \frac{d}{2}-1 \end{pmatrix}\begin{pmatrix} j-1 \\ \frac{d}{2}-1 \end{pmatrix} \\
&\quad - \sum_{i=n+\frac{d}{2}}^{\infty}\sum_{j=n+\frac{d}{2}}^{\infty} 2^{-(p+1)(i+j)} \begin{pmatrix} i-1 \\ \frac{d}{2}-1 \end{pmatrix}\begin{pmatrix} j-1 \\ \frac{d}{2}-1 \end{pmatrix}
\end{aligned}$$

Applying BG Lemma 3.7 which shows $\sum_{i=n+d}^{\infty} 2^{-si}\begin{pmatrix} i-1 \\ d-1 \end{pmatrix} \leq 2^{-sn}\cdot 2^{-sd}$. $2\cdot A(d,n) := C(s,d,n)$ where $A(d,n) = \sum_{k=0}^{d-1}\begin{pmatrix} n+d+i-1 \\ d-1 \end{pmatrix} = \frac{n^{d-1}}{(d-1)!} + O(n^{d-2})$, obtain upper bound

$$\begin{aligned}
\sum_{\max\{|\ell^{d_1}|_1,|\ell^{d_2}|_1\} > n+\frac{d}{2}-1} 2^{-|\ell\cdot(p+1)|_1} &\leq C(p+1,\frac{d}{2},n)\cdot \\
&\quad (2\sum_{j=\frac{d}{2}}^{\infty} 2^{-(p+1)j}\begin{pmatrix} j-1 \\ \frac{d}{2}-1 \end{pmatrix} - \sum_{j=n+\frac{d}{2}}^{\infty} 2^{-(p+1)j}\begin{pmatrix} j-1 \\ \frac{d}{2}-1 \end{pmatrix})
\end{aligned}$$

Elementary but tedious combinatorics shows that

$$\sum_{j=\frac{d}{2}}^{\infty} 2^{-(p+1)j} \left(\begin{array}{c} j-1 \\ \frac{d}{2}-1 \end{array} \right) \quad = \quad 2^{-(p+1)} \sum_{j=\frac{d}{2}}^{\infty} 2^{-(p+1)(j-1)} \left(\begin{array}{c} j-1 \\ \frac{d}{2}-1 \end{array} \right) = (\frac{2^{p+1}}{2^{p+1}-1})^{\frac{d}{2}}$$

and so the first series in the parentheses converges to a constant not depending on n and the second goes to 0 in n, producing upper bound

$$\begin{aligned} \sum_{\max\{|\ell^{d_1}|_1, |\ell^{d_2}|_1\} > n+\frac{d}{2}-1} 2^{-|\ell \cdot (p+1)|_1} &\leq 2(\frac{2^{p+1}}{2^{p+1}-1})^{\frac{d}{2}} C(p+1, \frac{d}{2}, n) \\ &\leq 2^{-(p+1)n} \cdot 4(\frac{1}{2^{p+1}-1})^{\frac{d}{2}} \cdot A(\frac{d}{2}, n) \\ &\leq 2^{-(p+1)n} \cdot 4(\frac{1}{2^{p+1}-1})^{\frac{d}{2}} \cdot (\frac{n^{\frac{d}{2}-1}}{(\frac{d}{2}-1)!} + O(n^{\frac{d}{2}-2})) \end{aligned}$$

Applying this to inequality (34), obtain bound

$$\left\| u - u_n^{p,\{1,1\}} \right\|_\infty \quad \leq \quad 0.5585^d c(p) |u|_{p+1,\infty} 2^{-(p+1)n} \cdot 4(\frac{1}{2^{p+1}-1})^{\frac{d}{2}} \cdot A(\frac{d}{2}, n) + + O(2^{-(p+1)n})$$

which up to constants is order $O(2^{-(p+1)n} n^{\frac{d}{2}-1})$

To find the number of function evaluations used, by BG Lemma (3.6), the number of points used in a d-dimensional sparse grid is $|V_n^{(1)}| = \sum_{i=0}^{n-1} 2^i \cdot \left(\begin{array}{c} d-1+i \\ d-1 \end{array} \right) = 2^n \cdot \left(\frac{n^{d-1}}{(d-1)!} + O(n^{d-2}) \right)$, so a tensor product of two such constructions in dimension $\frac{d}{2}$ each uses

$$N = \left(\sum_{i=0}^{n-1} 2^i \cdot \left(\begin{array}{c} \frac{d}{2}-1+i \\ \frac{d}{2}-1 \end{array} \right) \right)^2 = 2^{2n} \cdot \left(\frac{n^{d-2}}{((\frac{d}{2}-1)!)^2} + O(n^{d-3}) \right)$$

points to achieve this error. So, using $\log_2 N \approx 2n + (d-2)\log_2(n)$ and $\epsilon = O(2^{-(p+1)n} n^{\frac{d}{2}-1})$, obtain

$$\epsilon = O(N^{-(p+1)/2} n^{(p+1)(\frac{d}{2}-1)} n^{\frac{d}{2}-1}) = O(N^{-(p+1)/2} \, |\log_2 N|^{(p+2)(\frac{d}{2}-1)})$$

as claimed. Inverting to find order gives

$$N = O(\epsilon^{\frac{-2}{p+1}} \, |\log_2 \epsilon|^{\frac{(p+2)(d-2)}{p+1}})$$

□

Proof of Lemma (9')

Proof. I apply the same inductive argument as in Lemma (9), with respect to the new set of norms. I show only the case for norms of type a, as the case for

norms of type $b$is entirely symmetric. For each ℓ, for each $j = 1 \dots 2d_2 + o^\ell$, the proof of inequality (33) can be performed inductively along the directed graph starting at input nodes $p^\ell = 1 \dots 2d_2 + o^\ell$ for which $\left\| \tilde{n}_j^{p^\ell} - n_j^{p^\ell} \right\|_{p_j^\ell, a} = \left\| \tilde{n}_j^{p^\ell m} - n_j^{p^\ell m} \right\|_{p_j^{\ell m}, a} = \left\| 1\{p^\ell = j\} - 1\{p^\ell = j\} \right\|_{p_j^{\ell m}, a} = 0$ as $M_j^{p^\ell} = \{1\}$ for any node $p^\ell \in \{1 \dots 2d_2 + o^\ell\}$, and $\tilde{n}_j^{p^\ell m} = 1\{p^\ell = j\} \in \hat{\mathcal{G}}_d^{[p_j^{\ell m}];[p_{ej}^{\ell m}]}$ for all $m \in M_j^{p^\ell}$, so the base case holds. For each p^ℓ, assume that inequality (33) holds for all $p \in dp(j, p^\ell)/p^\ell$, and that $\forall q \in q_j(p^\ell)$, $\forall m' \in M_j^q$, $\tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \in \hat{\mathcal{G}}_d^{[q_j^{m'}];[q_{ej}^{m'}]}$. Then, in the case where node p^ℓ has identity for a linear component, obtain bound $\left\| \tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')} \right\|_{p_j^{\ell m}, a}$

$$
\begin{aligned}
&= \Bigg\| \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \cdot \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \\
&\quad - \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \cdot n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \Bigg\|_{p_j^{\ell m}, a} \\
&\leq \left\| \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) - \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \right\|_{p_j^{\ell m}, a} \left\| \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right\|_\infty \\
&\quad + \left\| \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p \in par(p^\ell)}) \right\|_\infty \left\| \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) - n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right\|_{p_j^{\ell m}, a} \\
&\leq C \Big(\sum_{p \in par(p^\ell)} \left\| \tilde{n}^{*p}(s_{[p]}) - n^{*p}(s_{[p]}) \right\|_\infty^{\alpha_d^{p^\ell(p)}} + c^a_{p^\ell(p)} + \sum_{(p,\tilde{m}) \in dp(j,q,m')} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{s \in par(p)} (\upsilon^p)^{\alpha_d^{p(s)}} + c^a_{p(s)}) \Big) \\
&\leq (\zeta_d^{p_{dj}^\ell}(K_{[p_{dj}^\ell]}) \cdot 0 + C \Big(\sum_{p \in par(p^\ell)} (\upsilon^p)^{\alpha_d^{p^\ell(p)}} + c^a_{p^\ell(p)} + \sum_{(p,\tilde{m}) \in dp(j,q,m')} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{s \in par(p)} (\upsilon^p)^{\alpha_d^{p(s)}} + c^a_{p(s)}) \Big) \\
&= C \sum_{(p,\tilde{m}) \in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q \in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c^a_{p(q)})
\end{aligned}
$$

The justifications for these steps are identical to those in Lemma (9) except that in the first inequality I use that $\|f(.)g(.)\|_{p_j^{\ell m}, a} \leq \|f(.)\|_{p_j^{\ell m}, a} \|g(.)\|_\infty$ by the definition of this norm and and in the second inequality I use that all functions and their approximations are assumed bounded, and that the $\|.\|_{p_j^{\ell m}, a}$ norm is bounded using Hölder continuity, the triangle inequality, commutativity of the supremum with respect to monotonic functions, and finite support.

In the case where node p^ℓ has an integral operator for a linear component,

obtain bound $\left\|\tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')}\right\|_{p_j^{\ell m},a}$

$$
\begin{aligned}
=&\left\| \sum_{i=1}^{K_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]}} \pi_i^{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} k^{p^\ell}(s_{[p_a^\ell],i}) \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell],i}, \{\tilde{n}^{*p}(s_{[p],i})\}_{p\in par(p^\ell)}) \cdot \tilde{n}_j^{qm'}(s_{[q_j^{m'}]i}, s'_{[q_{ej}^{m'}]}) \right. \\
&\left. - \int k^{p^\ell}(s_{[p_a^\ell]}) \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \cdot n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) ds_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} \right\|_{p_j^{\ell m},a} \\
\leq& \zeta_{dK}^{p^\ell jm} + \left\| \int k^{p^\ell}(s_{[p_a^\ell]}) \left[\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \cdot \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right.\right. \\
&\left.\left. - \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \cdot n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right] ds_{([p_b^\ell]\cap[q_{ej}^{m'}])\cup[p_{dj}^\ell]} \right\|_{p_j^{\ell m},a} \\
\leq& \zeta_{dK}^{p^\ell jm} + C \left\| k^{p^\ell}(s_{[p_a^\ell]}) \left[\frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{\tilde{n}^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \right.\right. \\
&\left.\left. - \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \right] \right\|_{p_j^{\ell m},a} \left\| \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right\|_\infty \\
&+ C \left\| k^{p^\ell}(s_{[p_a^\ell]}) \frac{\partial}{dn^q} f^{p^\ell}(s_{[p_I^\ell]}, \{n^{*p}(s_{[p]})\}_{p\in par(p^\ell)}) \right\|_\infty \left\| \tilde{n}_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) - n_j^{qm'}(s_{[q_j^{m'}]}, s'_{[q_{ej}^{m'}]}) \right\|_{p_j^{\ell m},a} \\
\leq& \zeta_{dK}^{p^\ell jm} + C(\sum_{p\in par(p^\ell)} \left\| \tilde{n}^{*p}(s_{[p]}) - n^{*p}(s_{[p]}) \right\|_\infty^{\alpha_d^{p(q)}} + \sum_{(p,\tilde{m})\in dp(j,q,m')} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{s\in par(p)} (\upsilon^p)^{\alpha_d^{p(s)}} + c_{p(s)}^a)) \\
\leq& (\zeta_{dK}^{p^\ell jm} \cdot 1 + C(\sum_{p\in par(p^\ell)} (\upsilon^p)^{\alpha_d^{p^\ell(p)}} + c_{p^\ell(p)}^a + \sum_{(p,\tilde{m})\in dp(j,q,m')} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{s\in par(p)} (\upsilon^p)^{\alpha_d^{p(s)}} + c_{p(s)}^a)) \\
=& C \sum_{(p,\tilde{m})\in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q\in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a)
\end{aligned}
$$

where the analysis is equivalent to that in Lemma (9) with modifications as above, except that in the first inequality I have applied condition **??**(iv)(c).

Applying this inequality to the definition of $\tilde{n}_j^{p^\ell}$, obtain $\left\|\tilde{n}_j^{p^\ell} - n_j^{p^\ell}\right\|_{p_j^\ell,a}$

$$
\begin{aligned}
&= \left\| \sum_{m\in M_j^{p^\ell}} \tilde{n}_j^{p^\ell m(q,m')} - \sum_{m\in M_j^{p^\ell}} n_j^{p^\ell m(q,m')} \right\|_{p_j^\ell,a} \\
&\leq \sum_{m\in M_j^{p^\ell}} \left\| \tilde{n}_j^{p^\ell m(q,m')} - n_j^{p^\ell m(q,m')} \right\|_{p_j^{\ell m},a} \\
&\leq C \sum_{m\in M_j^{p^\ell}} \sum_{(p,\tilde{m})\in dp(j,p^\ell,m)} (\zeta_{dK}^{pj\tilde{m}} 1_p + \sum_{q\in par(p)} (\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a)
\end{aligned}
$$

Next $\tilde{n}_j^{p^\ell m(q,m')} \in \hat{\mathcal{G}}_d^{[p_j^{\ell m}];[p_{ej}^{\ell m}]}$ in either of the above cases, so the inductive

hypothesis holds and the claim is true for all p^ℓ. Symmetric computations prove an identical bound for norm $\|.\|_{p_j^\ell,b}$.

Next, note that by the construction of the graph, the final node for each ℓ contains no intermediate variables and so $\|.\|_{P_j^\ell,a} = \|.\|_{1/\infty}$ and $\|.\|_{P_j^\ell,b} = \|.\|_{\infty/1}$. So, following Lemma (9) for the final node I have in the case where ℓ satisfies condition (4)(b)(ii)

$$\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}(M_{[\ell^o]}\mathcal{F}^{\ell}_{\overrightarrow{g}_j(in)}(\Pi^{[j]})^{-1}M^*_{[j]})_{ik}\phi_i\phi_k - n_j^{P^\ell}\right\|_{1/\infty} =$$

$$\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}((M_{[\ell^o]}\otimes M_{[j]})\{\{\tilde{n}_j^{P^\ell}(s_{[\ell^o],r_1},s'_{[j],r_2})\}_{r_2=1}^{K_{[j]}}\}_{r_1=1}^{K_{[\ell^o]}})_{ik}\phi_i\phi_k - n_j^{P^\ell}\right\|_{1/\infty} \leq$$

$$\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}((M_{[\ell^o]}\otimes M_{[j]})\{\{\tilde{n}_j^{P^\ell}(s_{[\ell^o],r_1},s'_{[j],r_2})\}_{r_2=1}^{K_{[j]}}\}_{r_1=1}^{K_{[\ell^o]}})_{ik}\phi_i\phi_k - \tilde{n}_j^{P^\ell}\right\|_{1/\infty} + \left\|\tilde{n}_j^{P^\ell} - n_j^{P^\ell}\right\|_{1/\infty}$$

$$\leq \sum_{m\in M_j^{P^\ell}}\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}((M_{[\ell^o]}\otimes M_{[j]})\{\{\tilde{n}_j^{P^\ell m}(s_{[P_j^{\ell m}],r_1},s'_{[p_{ej}^{\ell m}],r_2})\}_{r_2=1}^{K_{[j]}}\}_{r_1=1}^{K_{[\ell^o]}})_{ik}\phi_i\phi_k - \tilde{n}_j^{P^\ell}\right\|_{1/\infty} + \left\|\tilde{n}_j^{P^\ell} - n_j^{P^\ell}\right\|_{1/\infty}$$

$$\leq C\sum_{m\in M_j^{P^\ell}}(\zeta_a^{\ell jm}(K_{[\chi+1]\times[j]}) + \sum_{(p,\tilde{m})\in dp(j,P^\ell,m)}(\zeta_{dK}^{pj\tilde{m}}1_p + \sum_{q\in par(p)}(\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a))$$

and in the case where ℓ satisfies condition (4)(b)(i)

$$\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}(M_{[\ell^o]}(\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(p^\ell(out))}})^{-1}\mathcal{F}^{\ell}_{\overrightarrow{g}_{j(in)}}(\Pi^{[j]})^{-1}M^*_{[j]})_{ik}\phi_i\phi_k - n_j^{P^\ell}/n_{j(out)}^{P^\ell}\right\|_{1/\infty} =$$

$$\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}((M_{[\ell^o]}\otimes M_{[j]})\{\{\tilde{n}_j^{P^\ell}(s_{[\ell^o],r_1},s'_{[j],r_2})/\tilde{n}_{j(out)}^{P^\ell}(s_{[\ell^o],r_1})\}_{r_2=1}^{K_{[j]}}\}_{r_1=1}^{K_{[\ell^o]}})_{ik}\phi_i\phi_k - n_j^{P^\ell}/n_{j(out)}^{P^\ell}\right\|_{1/\infty} \leq$$

$$\sum_{m\in M_j^{P^\ell}}\left\|\sum_{k=1}^{K_{[\ell^o]}}\sum_{i=1}^{K_{[j]}}((M_{[\ell^o]}\otimes M_{[j]})\{\{\tilde{n}_j^{P^\ell m}(s_{[\ell^o],r_1},s'_{[j],r_2})/\tilde{n}_{j(out)}^{P^\ell}(s_{[\ell^o],r_1})\}_{r_2=1}^{K_{[j]}}\}_{r_1=1}^{K_{[\ell^o]}})_{ik}\phi_i\phi_k - \tilde{n}_j^{P^\ell m}/\tilde{n}_{j(out)}^{P^\ell}\right\|_{1/\infty} +$$

$$\left\|\tilde{n}_j^{P^\ell}/\tilde{n}_{j(out)}^{P^\ell} - n_j^{P^\ell}/n_{j(out)}^{P^\ell}\right\|_{1/\infty} \leq$$

$$C\sum_{m\in M_j^{P^\ell}}(\zeta_d^{\ell jm}(K_{[\chi+1]\times[j]}) + \left\|\tilde{n}_j^{P^\ell} - n_j^{P^\ell}\right\|_{1/\infty} + \left\|\tilde{n}_{j(out)}^{P^\ell} - n_{j(out)}^{P^\ell}\right\|_{1/\infty}) \leq$$

$$C(\sum_{m\in M_j^{P^\ell}}\zeta_{da}^{\ell jm}(K_{[\chi+1]\times[j]}) + \sum_{m\in M_j^{P^\ell}}\sum_{(p,\tilde{m})\in dp(j,P^\ell,m)}(\zeta_{dK}^{pj\tilde{m}}1_p + \sum_{q\in par(p)}(\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a) +$$

$$\sum_{m\in M_{j(out)}^{P^\ell}}\sum_{(p,\tilde{m})\in dp(j(out),P^\ell,m)}(\zeta_{dK}^{pj(out)\tilde{m}}1_p + \sum_{q\in par(p)}(\upsilon^p)^{\alpha_d^{p(q)}} + c_{p(q)}^a))$$

as before, symmetric bounds apply for the $\|.\|_{\infty/1}$ norm. □

Proof of Lemma (21)

Proof. Let $f(x_1,\ldots,x_{d_x},y_1,\ldots,y_{d_y})\in\mathcal{PR}_{0,1}^{\alpha,N}(\mathcal{X},\mathcal{Y})$, Letting i and j be multi-indices, $K_x = K_z^{d_x}$, $K_y = K_z^{d_y}$, $K = K_xK_y$, $[\vec{f}]_{ij} = \frac{1}{\sqrt{K}}f(x_{i_1},\ldots,x_{i_{d_x}},y_{j_1},\ldots,y_{j_{d_y}})$ and $\varphi_i(x) = \prod_{z=1\ldots d_x}\phi_{i_z}(x_z)$, $\varphi_j(y) = \prod_{z=1\ldots d_y}\phi_{j_z}(y_z)$. $\hat{f} = \sum_{i=1}^{K_x}\sum_{j=1}^{K_y}[\vec{f}]_{ij}\varphi_i(.)\varphi_j(.)$ is the histogram approximation of f. The $\|.\|_{1/\infty}$ norm accuracy of this approximation is given by

$$\begin{aligned}\left\|\hat{f}(x,y) - f(x,y)\right\|_{1/\infty} &\leq \left\|\hat{f}(x,y) - \frac{1}{\sqrt{K_x}}\sum_{i=1}^{K_x}f(x_i,y)\varphi_i(x)\right\|_{1/\infty} + \left\|\frac{1}{\sqrt{K_x}}\sum_{i=1}^{K_x}f(x_i,y)\varphi_i(x) - f(x,y)\right\|_{1/\infty} \\ &= (a)+(b)\end{aligned}$$

where (a) is the error in y and (b) is the error in approximation of x. Let (I) refer to the set of scaling functions $\varphi_i(x)$ whose domain intersects with ∂B_n for some n, and (II) refer to the set of "interior" scaling functions which do not intersect any of these boundaries. By disjoint support of the scaling functions,

obtain

$$\begin{aligned}(b) &\leq \left\| \frac{1}{\sqrt{K_x}} \sum_{i\in(I)} (f(x_i,y)\varphi_i(x) - f(x,y)) 1\{x \in supp\varphi_i\} \right\|_{1/\infty} \\ &\quad + \left\| \frac{1}{\sqrt{K_x}} \sum_{i\in(II)} (f(x_i,y)\varphi_i(x) - f(x,y)) 1\{x \in supp\varphi_i\} \right\|_{1/\infty} \\ &= (b)_I + (b)_{II}\end{aligned}$$

Each scaling function φ_i has diameter $\sqrt{d_x}K_z^{-1}$, while the volume of all such domains which intersect with some ∂B_n is less than the volume of a tube of radius $\sqrt{d_x}K_z^{-1}$ around ∂B_n which is, following Grohs et al. [2018] Lemma 2 and the assumptions made on ∂B_n, less than $C\sqrt{d_x}K_z^{-1}$ for some C. As there are at most N such regions, and $|f(x_i,y)\varphi_i(x) - f(x,y)|$ is bounded, obtain

$$(b)_I \leq CN\sqrt{d_x}K_z^{-1}$$

for some C. As outside of these boundary regions, $f(x,y)$ is α-Hölder in x and the L^1 norm is dominated by the L^∞ norm, calculations identical to those of Lemma (17) imply

$$(b)_{II} \leq C\left(\sqrt{d_x}K_z^{-1}\right)^{\min\{\alpha,1\}}$$

To bound (a), apply the steps of Lemma (17) to obtain the bound, for some $s \in \mathcal{Y}$ corresponding to the scaling function φ_j with the largest error

$$(a) \leq \sup_{u\in[\frac{-1}{2K_z},\frac{1}{2K_z})^{d_y}} \int |f(x,s+u) - f(x,s))|\, dx$$

Applying Taylor's theorem with integral form of the remainder and weak differentiability of f with respect to y, this is bounded by

$$\sup_{u\in[\frac{-1}{2K_z},\frac{1}{2K_z})^{d_y}} \int \left| \int_0^1 \frac{\partial}{dy} f(x,s+tu) dt \right| dx$$

within any region $\{B_n\}_{n=1}^N$ or $B_0 := \mathcal{Y}/\cup_{n=1}^N B_n$, $\frac{\partial}{dy}f(x,y)$ is α-Hölder, and so between any 2 points $s+t_0u$ and $s+t_1u$ in the same region B_n $\left|\frac{\partial}{dy}f(x,s+t_1u) - \frac{\partial}{dy}f(x,s+t_0u)\right| \leq CL((t_1-t_0)(\sqrt{d_y}\frac{K_z}{2}^{-1})^{\min\{\alpha,1\}}$ and so the integral of all points in this region is less than $\int_0^1 L(t\sqrt{d_y}\frac{K_z}{2}^{-1})^{\min\{\alpha,1\}}dt = L\frac{1}{\alpha+1}(\sqrt{d_y}\frac{K_z}{2})^{-\min\{\alpha,1\}}$. As the $N+1$ regions form a partition of the hypercube, the above may be bounded by

$$(N+1)CL\frac{1}{\alpha+1}(\sqrt{d_y}\frac{K_z}{2})^{-\min\{\alpha,1\}}$$

Combining terms, obtain bound

$$\begin{aligned}\left\|\hat{f}(x,y)-f(x,y)\right\|_{1/\infty} &\leq C((N+1)(\sqrt{d_y}\frac{K_z}{2})^{-\min\{\alpha,1\}}+\left(\sqrt{d_x}K_z^{-1}\right)^{\min\{\alpha,1\}}+N\sqrt{d_x}K_z^{-1}) \\ &= O({K_z}^{-\min\{\alpha,1\}})=O(K^{-\frac{\min\{\alpha,1\}}{d}})\end{aligned}$$

as claimed. To obtain $\left\|\hat{f}(x,y)-f(x,y)\right\|_{\infty/1}=O(1)$, simply note that f and its approximation are bounded over a bounded domain.

The results for $f\in\mathcal{PR}_{1,0}^{\alpha,N}(\mathcal{X},\mathcal{Y})$ follow by symmetry. □

Proof. of Lemma (24). Follows closely the proof of Lemma (17), with a small number of additional steps. Let $f(x_1,\ldots,x_{d_x},y_1,\ldots,y_{d_y})\in\mathcal{F}$, letting i be a (pair of) multi-indices $(i^x,i^y)=(i_1^x,\ldots,i_{d_x}^x,i_1^y,\ldots,i_{d_y}^y)$, and $K=K_z^{d_x+d_y}$, $[\vec{f}]_i=\frac{1}{\sqrt{K}}f(x_{i^x},y_{i^y})$ and $\varphi_i(x,y)=\tilde{\phi}_{i^x}(x)\tilde{\phi}_{i^y}(y)$. $\hat{f}=\sum_{i=1}^K[\vec{f}]_i\varphi_i(.)$ is the transformed histogram approximation of f. The sup norm accuracy of this approximation is given by

$$\begin{aligned}\sup_{x,y\in\mathcal{X}\times\mathcal{Y}}\left|\hat{f}(x,y)-f(x,y)\right| &\leq \sup_{x,y\in\mathcal{X}\times\mathcal{Y}}\left|\hat{f}(x,y)-\sum_{i=1}^K\langle f,\varphi_i\rangle\varphi_i(x,y)\right| \\ &\quad+\sup_{x,y\in\mathcal{X}\times\mathcal{Y}}\left|\sum_{i=1}^K\langle f,\varphi_i\rangle\varphi_i(x,y)-f(x,y)\right| \\ &= (I)+(II)\end{aligned}$$

where (I) is the quadrature error and (II) is the projection error.

Letting $j=(j^x,j^y)$ denote the multi-index of the region in which the supremum of part (II) is achieved and $s=(s_1,s_2)$ denote any point in that region

$$\begin{aligned}(II) &= |\langle f,\varphi_j\rangle\varphi_j(s)-f(s)| \\ &= (c_{j^x}^x)^{-\frac{1}{2}}(c_{j^y}^y)^{-\frac{1}{2}}\int\int\varphi_j(x,y)(f(x,y)-f(s_1,s_2))dxdy \\ &= (c_{j^x}^x)^{-\frac{1}{2}}(c_{j^y}^y)^{-\frac{1}{2}}\int\int\varphi_j(s_1+u_1,s_1+u_2)(f(s_1+u_1,s_2+u_2)-f(s_1,s_2))du_1du_2 \\ &\leq (c_{j^x}^x)^{-\frac{1}{2}}(c_{j^y}^y)^{-\frac{1}{2}}\int\int|\varphi_j(s_1+u_1,s_2+u_2)|\,|f(s_1+u_1,s_2+u_2)-f(s_1,s_2)|\,du_1du_2 \\ &\leq \sup_{(\tau_x^{-1,K}(s_1+u_1),\tau_y^{-1,K}(s_2+u_2))\in\text{support }\varphi_j(.)}|f(s_1+u_1,s_2+u_2)-f(s_1,s_2)|\cdot \\ &\quad (c_{j^x}^x)^{-\frac{1}{2}}(c_{j^y}^y)^{-\frac{1}{2}}\int\int|\varphi_j(s_1+u_1,s_2+u_2)|\,du_1du_2 \\ &= \sup_{(\tau_x^{-1,K}(s_1+u_1),\tau_y^{-1,K}(s_2+u_2))\in\text{support }\varphi_j(.)}|f(s_1+u_1,s_2+u_2)-f(s_1,s_2)| \\ &\leq L(\frac{2C+\sqrt{d_x}}{K_z})^{\min\{\alpha,1\}}+L(\frac{2C+\sqrt{d_y}}{K_z})^{\min\{\alpha,1\}}=O(K_z^{-\min\{\alpha,1\}})\end{aligned}$$

Where the second and sixth lines follow by the normalization of φ_j from Condition 22(iii), and the last by applying the definition of piecewise regular functions and by Condition 22(i)(a), in the region over which the difference is taken, f is uniformly α-Hölder in arguments x and y, and by condition 22(ii) the maximum distance between two points in the set $\{u_1 :\ \tau_x^{-1,K}(s_1+u_1) \in \text{support } \tilde{\phi}_{j^x}(x)\}$ is no more than $\frac{\sqrt{d_x}}{K_z} + \frac{2C}{K_z}$ and in the set $\{u_2 :\ \tau_x^{-1,K}(s_2+u_2) \in \text{support } \tilde{\phi}_{j^y}(y)\}$ is no more than $\frac{\sqrt{d_y}}{K_z} + \frac{2C}{K_z}$.

The bound on (I) follows from

$$\begin{aligned}
(I) &= \sup_j \sup_{x,y\in\mathcal{X}\times\mathcal{Y}} \left|([\vec{f}]_j - \langle f,\varphi_j\rangle)\varphi_j(x,y)\right| \\
&= \sup_j \sup_{x,y\in\mathcal{X}\times\mathcal{Y}} \left|(c_{j^x}^x)^{-\frac{1}{2}}(c_{j^y}^y)^{-\frac{1}{2}}\frac{1}{\sqrt{K}} f(x_{j^x},y_{j^y}) - \langle f,\varphi_j\rangle\,\varphi_j(x,y)\right| \\
&= \sup_j \sup_{x,y\in\mathcal{X}\times\mathcal{Y}} \left|\left(\frac{1/K_z^{d_x}}{c_{j^x}^x}\right)^{\frac{1}{2}}\left(\frac{1/K_z^{d_y}}{c_{j^y}^y}\right)^{\frac{1}{2}} f(x_{j^x},y_{j^y}) - \langle f,\varphi_j\rangle\,\varphi_j(x,y)\right| \\
&\leq \sup_j \sup_{x,y\in\mathcal{X}\times\mathcal{Y}} \left|\left(\frac{1/K_z^{d_x}}{c_{j^x}^x}\right)^{\frac{1}{2}}\left(\frac{1/K_z^{d_y}}{c_{j^y}^y}\right)^{\frac{1}{2}} f(x_{j^x},y_{j^y}) - f(x_{j^x},y_{j^y})\right| \\
&\quad + \sup_j \sup_{x,y\in\mathcal{X}\times\mathcal{Y}} |f(x_{j^x},y_{j^y}) - \langle f,\varphi_j\rangle\,\varphi_j(x,y)| \\
&= (a)+(b)
\end{aligned}$$

□

Here, $(b) = O(K_z^{-\min\{\alpha,1\}})$ by the same steps as in the bound of (II) and $(a) \leq O(K_z^{-\frac{1}{2}})$ by boundedness of f and Condition 22(iii). Combining these, the error from part (I) dominates if $\alpha > \frac{1}{2}$, otherwise both terms contribute at the same rate, and so the approximation scheme is sup norm accurate at rate

$$O(K_z^{-\min\{\alpha,\frac{1}{2}\}})$$

Proof. of Lemma (25).

By the triangle inequality $\sup_{\{f\in\Lambda^{\alpha_y}:\ \|f\|_{L^2(\mathcal{Y})}=1\}} \left\|\sum_{i=1}^{K_x}\sum_{j=1}^{K_y} Z_{ij}^K \langle\phi_j(y),f(.)\rangle\,\phi_i(x) - Z[f(.)]\right\|_{L^2(\mathcal{X})}$

$$\begin{aligned}
&\leq \left\|\sum_{i=1}^{K_x}\sum_{j=1}^{K_y} Z_{ij}^K \left\langle\tilde{\phi}_j(y),[.]\right\rangle\tilde{\phi}_i(x) - Z[.]\right\|_{op} + \\
&\quad \sup_{\{f\in\Lambda^{\alpha_y}:\ \|f\|_{L^2(\mathcal{Y})}=1\}} \left\|\sum_{i=1}^{K_x}\sum_{j=1}^{K_y} Z_{ij}^K \langle\phi_j(y),f(.)\rangle\,\phi_i(x) - \sum_{i=1}^{K_x}\sum_{j=1}^{K_y} Z_{ij}^K \left\langle\tilde{\phi}_j(y),f(.)\right\rangle\tilde{\phi}_i(x)\right\|_{L^2(\mathcal{X})} \\
&\leq \epsilon_K + (A)
\end{aligned}$$

and

$$\begin{aligned}
(A) &= \sup_{\{f\in\Lambda^{\alpha_y}:\ \|f\|_{L^2(\mathcal{Y})}=1\}} \left\| \sum_{i=1}^{K_x}\sum_{j=1}^{K_y} Z_{ij}^K \left(\langle \phi_j(y), f(.)\rangle \phi_i(x) - \left\langle \tilde{\phi}_j(y), f(.)\right\rangle \tilde{\phi}_i(x)\right)\right\|_{L^2(\mathcal{X})} \\
&\leq \sup_{\{f\in\Lambda^{\alpha_y}:\ \|f\|_{L^2(\mathcal{Y})}=1\}} \left\| \sum_{i=1}^{K_x}\sum_{j=1}^{K_y} Z_{ij}^K \left\langle \tilde{\phi}_j(y), f(.)\right\rangle \left(\phi_i(x) - \tilde{\phi}_i(x)\right)\right\|_{L^2(\mathcal{X})} + \\
&\quad \sup_{\{f\in\Lambda^{\alpha_y}:\ \|f\|_{L^2(\mathcal{Y})}=1\}} \left\| \sum_{i=1}^{K_x}\sum_{j=1}^{K_y} Z_{ij}^K \phi_i(x) \left\langle \phi_j(y) - \tilde{\phi}_j(y), f(.)\right\rangle \right\|_{L^2(\mathcal{X})} \\
&= (A1) + (A2)
\end{aligned}$$

Using the fact that projections are isometries in L^2 and defining $\{e_j\}_{j=1}^{K_y}$ as the canonical basis vectors in Euclidean space, $(A2)$ has upper bound

$$\left\|Z^K\right\|_{op} \cdot \sup_{\{f\in\Lambda^{\alpha_y}:\ \|f\|_{L^2(\mathcal{Y})}=1\}} \left\| \sum_{j=1}^{K_y} \left\langle \phi_j(y) - \tilde{\phi}_j(y), f(.)\right\rangle e_j \right\|_2$$

For any f in $\{f\in\Lambda^{\alpha_y}:\ \|f\|_{L^2(\mathcal{Y})}=1\}$

$$\begin{aligned}
\left\| \sum_{j=1}^{K_y} \left\langle \phi_j(y) - \tilde{\phi}_j(y), f(.)\right\rangle e_j \right\|_2 &= \left(\sum_{j=1}^{K_y} \left(\int (\phi_j(y) - \tilde{\phi}_j(y)) f(y) dy \right)^2 \right)^{\frac{1}{2}} \\
&\leq \left(\sum_{j=1}^{K_y} \left(\int \phi_j(y)(f(y) - f(\tau_y^K(y)))dy \right)^2 \right)^{\frac{1}{2}} \\
&\quad + \left(\sum_{j=1}^{K_y} \left(\int \phi_j(y) f(\tau_y^K(y)) dy - \int \tilde{\phi}_j(y) f(y) dy \right)^2 \right)^{\frac{1}{2}} \\
&= (A2a) + (A2b)
\end{aligned}$$

I have that

$$\begin{aligned}
(A2a) &\leq C \left\|\tau(y) - y\right\|_\infty^{\alpha_y} \\
&\leq C K_z^{-\alpha_y}
\end{aligned}$$

for some C, using the Hölder condition on $f(.)$, the fact that projection is an isometry and sup norm dominates L^2 norm, and Condition 22(ii).

$$
\begin{aligned}
(A2b) &= \left(\sum_{j=1}^{K_y}\left(K_z^{\frac{d_y}{2}}\int 1\{u\in \text{supp }\phi_j(.)\}f(\tau_y^K(u))du-(c_j^y)^{-\frac{1}{2}}\int 1\{\tau_y^{-1,K}(y)\in \text{supp }\phi_j(.)\}f(y)dy\right)^2\right)^{\frac{1}{2}} \\
&= \left(\sum_{j=1}^{K_y}\left(K_z^{\frac{d_y}{2}}\int 1\{u\in \text{supp }\phi_j(.)\}f(\tau_y^K(u))du-\right.\right. \\
&\quad \left.\left.(c_j^y)^{-\frac{1}{2}}\int 1\{u\in \text{supp }\phi_j(.)\}f(\tau_y^K(u)))\left(\det\frac{\partial}{du}\tau_y^{-1,K}(\tau_y^K(u))\right)^{-1}du\right)^2\right)^{\frac{1}{2}} \\
&= \left(\sum_{j=1}^{K_y}\left(\int 1\{u\in \text{supp }\phi_j(.)\}f(\tau_y^K(u))\left(K_z^{\frac{d_y}{2}}-(c_j^y)^{-\frac{1}{2}}\left(\det\frac{\partial}{du}\tau_y^{-1,K}(\tau_y^K(u))\right)^{-1}\right)du\right)^2\right)^{\frac{1}{2}} \\
&\leq \left(\sum_{j=1}^{K_y}\left(\int 1\{u\in \text{supp }\phi_j(.)\}f(\tau_y^K(u))^2du\right)\right)^{\frac{1}{2}}\cdot \\
&\quad \left(\sum_{j=1}^{K_y}\left(\int 1\{u\in \text{supp }\phi_j(.)\}\left(K_z^{\frac{d_y}{2}}-(c_j^y)^{-\frac{1}{2}}\left(\det\frac{\partial}{du}\tau_y^{-1,K}(\tau_y^K(u))\right)^{-1}\right)^2du\right)\right)^{\frac{1}{2}} \\
&= \left(\int f(\tau_y^K(u))^2du\right)^{\frac{1}{2}}\max_j\left|K_z^{\frac{d_y}{2}}-(c_j^y)^{-\frac{1}{2}}\right|\left(\int\left(\det\frac{\partial}{du}\tau_y^{-1,K}(\tau_y^K(u))\right)^{-2}du\right)^{\frac{1}{2}} \\
&\leq C\max_j\left|K_z^{\frac{d_y}{2}}-(c_j^y)^{-\frac{1}{2}}\right| \\
&= O(K_z^{-\frac{1}{2}})
\end{aligned}
$$

where the second line follows by a change of variables, the fourth by Cauchy-Schwarz (applied twice, once for the integral and once for the sum), the fifth by Condition 22(iii)(b), the boundedness of $f(.)$, and the last by Condition 22(iii)(a).

Together, along with the fact that $\left\|Z^K\right\|_{op}=\left\|\sum_{i=1}^{K_x}\sum_{j=1}^{K_y}Z_{ij}^K\left\langle\tilde{\phi}_j(y),[.]\right\rangle\tilde{\phi}_i(x)\right\|_{op}$ is bounded since projections are isometries and the latter is bounded because it converges to a bounded operator, these imply that

$$(A2)\leq O(K_z^{-\min\{\alpha_y,\frac{1}{2}\}})$$

To bound $(A1)$, note that by the bounded operator norm of $\left\|Z^K\right\|_{op}$ and by the fact that by the bounded L^2 norm of the input function f,

$$\sup_{\{f\in\Lambda^{\alpha_y}:\ \|f\|_{L^2(\mathcal{Y})}=1\}}\left\|\sum_{j=1}^{K_y}\left\langle\tilde{\phi}_j(y),f(.)\right\rangle e_j\right\|_2\leq 1$$

I have that for some constant C

$$
\begin{aligned}
(A1) \quad &\leq \sup_{\{v \in \mathbb{R}^{d_x} \colon \|v\|_2 \leq C\}} \left\| \sum_{i=1}^{K_x} v_i \left(\phi_i(x) - \tilde{\phi}_i(x) \right) \right\|_{L^2(\mathcal{X})} \\
&= \sup_{\{v \in \mathbb{R}^{d_x} \colon \|v\|_2 \leq C\}} \left(\int \left(\sum_{i=1}^{K_x} v_i \left(\sqrt{K_z^{d_x}} 1\{x \in \text{supp } \phi_i()\} - (c_i^x)^{-\frac{1}{2}} 1\{\tau_x^{-1,K}(x) \in \text{supp } \phi_i()\} \right) \right)^2 dx \right)^{\frac{1}{2}}
\end{aligned}
$$

By the fact that each hypercube defining the support of ϕ_i for some i is adjacent to no more than $3^{d_x} - 1$ other hypercubes in the partition and by Condition 22(ii), the support of $\tilde{\phi}_i$ can only overlap with that of these adjacent hypercubes, each term in the sum has non-zero product with at most a constant (independent of K_z) number of other terms. By Condition 22(ii) and (iii), each of these products has integral of order $\frac{1}{K_z}$, so the above term is bounded by

$$
\begin{aligned}
(A1) \quad &\leq \left(3^{d_x} - 1\right)^{\frac{1}{2}} \sup_{\{v \in \mathbb{R}^{d_x} \colon \|v\|_2 \leq C\}} \left(\sum_{i=1}^{K_x} v_i^2 \right)^{\frac{1}{2}} O(\frac{1}{\sqrt{K_z}}) \\
&= C \left(3^{d_x} - 1\right)^{\frac{1}{2}} O(\frac{1}{\sqrt{K_z}}) \\
&= O(\frac{1}{\sqrt{K_z}})
\end{aligned}
$$

Combining all terms yields the result. □

Proof. of Corollary (26).

By the assumptions of the corollary, the approximations of the policy operators with respect to the transformed histogram basis converges. Denoting these basis functions $\tilde{\varphi}_j(x)$ and $\tilde{\varphi}_i(y)$, this implies, denoting $(\tilde{h}_x^{\perp}, \tilde{g}_x^{\perp})$ the components of (h_x, g_x) orthogonal to the span of the K dimensional histogram basis constructed as in Algorithm (1) in Childers (2018), for $K \geq \bar{K}$,

$$
\left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[\tilde{h}_x^K \right]_{ij} \langle \tilde{\varphi}_j(x), (.) \rangle \, \tilde{\varphi}_i(y) + \tilde{h}_x^{\perp} - h_x[.] \right\|_{op} \leq \epsilon_K
$$

$$
\left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[\tilde{g}_x^K \right]_{ij} \langle \tilde{\varphi}_j(x), (.) \rangle \, \tilde{\varphi}_i(y) + \tilde{g}_x^{\perp} - g_x[.] \right\|_{op} \leq \epsilon_K
$$

As the conditions of Lemma (25) are verified, this implies

$$
\sup_{\{f \in \Lambda^{\alpha} \colon \|f\|_{\mathcal{H}_X} = 1\}} \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[\tilde{h}_x^K \right]_{ij} \langle \varphi_j(x), f(.) \rangle \, \varphi_i(x) - (h_x - \tilde{h}_x^{\perp})[f(.)] \right\|_{\mathcal{H}_X} \leq \epsilon_K + C K_z^{-\frac{1}{2}}
$$

$$\sup_{\{f\in\Lambda^\alpha:\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \sum_{i=1}^{K_x}\sum_{j=1}^{K_y} \left[\tilde{g}_x^K\right]_{ij} \langle \varphi_j(x), f(.)\rangle \, \varphi_i(y) - (g_x - \tilde{g}_x^{\perp})[f(.)] \right\|_{\mathcal{H}_Y} \leq \epsilon_K + CK_z^{-\frac{1}{2}}$$

Applying the triangle inequality, obtain

$$\begin{aligned}\sup_{\{f\in\Lambda^\alpha:\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \sum_{i=1}^{K_x}\sum_{j=1}^{K_y} \left[\tilde{h}_x^K\right]_{ij} \langle \varphi_j(x), f(.)\rangle \, \varphi_i(x) - h_x[f(.)] \right\|_{\mathcal{H}_X} &\leq \epsilon_K + CK_z^{-\frac{1}{2}} \\ &\quad + \sup_{\{f\in\Lambda^\alpha:\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \tilde{h}_x^{\perp}[f(.)] \right\|_{\mathcal{H}_X}\end{aligned}$$

By the boundedness of $\tilde{h}_x^{\perp}$,

$$\sup_{\{f\in\Lambda^\alpha:\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \tilde{h}_x^{\perp}[f(.)] \right\|_{\mathcal{H}_X} \leq C \sup_{\{f\in\Lambda^\alpha:\ \|f\|_{\mathcal{H}_X}=1\}} \left\| f(x) - \sum_{j=1}^{K_x} \langle \tilde{\varphi}_j(y), f(.)\rangle \, \tilde{\varphi}_j(x) \right\|_{\mathcal{H}_X}$$

As the norm is bounded by the sup norm, steps exactly analogous to the steps bounding Part (I) in the proof of Lemma (24) bound this by

$$O(K_z^{-\min\{\alpha,1\}}) \leq O(K_z^{-\frac{1}{2}})$$

and so

$$\sup_{\{f\in\Lambda^\alpha:\ \|f\|_{\mathcal{H}_X}=1\}} \left\| \sum_{i=1}^{K_x}\sum_{j=1}^{K_y} \left[\tilde{h}_x^K\right]_{ij} \langle \varphi_j(x), f(.)\rangle \, \varphi_i(x) - h_x[f(.)] \right\|_{\mathcal{H}_X} \leq \epsilon_K + CK_z^{-\frac{1}{2}}$$

The proof for g_x is symmetric.

For impulse response functions, for any finite m, note that the operator norm bound implies

$$\left\| \sum_{i=1}^{K_x}\sum_{j=1}^{K_y} \left[(\tilde{h}_x^K)^m\right]_{ij} \langle \tilde{\varphi}_j(x), f(.)\rangle \, \tilde{\varphi}_i(y) + (\tilde{h}_x^{\perp})^m - (h_x)^m[.] \right\|_{op}$$

applying the triangle inequality repeatedly

$$
\begin{aligned}
\leq \quad & \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[(\tilde{h}_x^K)^m \right]_{ij} \langle \tilde{\varphi}_j(x), (.) \rangle \, \tilde{\varphi}_i(y) + (\tilde{h}_x^{\perp})^m - \right. \\
& \left. \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[(\tilde{h}_x^K)^{m-1} \right]_{ij} \langle \tilde{\varphi}_j(x), h_x(.) \rangle \, \tilde{\varphi}_i(y) + (\tilde{h}_x^{\perp})^{m-1} h_x \right\|_{op} \\
& + \|h_x\|_{op} \left\| [\sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[(\tilde{h}_x^K)^{m-1} \right]_{ij} \langle \tilde{\varphi}_j(x), f(.) \rangle \, \tilde{\varphi}_i(y)] + (\tilde{h}_x^{\perp})^{m-1} - (h_x)^{m-1}[.] \right\|_{op} \\
\leq \quad & \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[(\tilde{h}_x^K)^{m-1} \right]_{ij} \langle \tilde{\varphi}_j(x), (.) \rangle \, \tilde{\varphi}_i(y) + (\tilde{h}_x^{\perp})^{m-1} \right\|_{op} \cdot \\
& \left\| \sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[(\tilde{h}_x^K) \right]_{ij} \langle \tilde{\varphi}_j(x), (.) \rangle \, \tilde{\varphi}_i(y) + (\tilde{h}_x^{\perp}) - h_x \right\|_{op} \\
& + \|h_x\|_{op} \left\| [\sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[(\tilde{h}_x^K)^{m-1} \right]_{ij} \langle \tilde{\varphi}_j(x), f(.) \rangle \, \tilde{\varphi}_i(y)] + (\tilde{h}_x^{\perp})^{m-1} - (h_x)^{m-1}[.] \right\|_{op} \\
\leq \quad & C_{m-1}(\epsilon_K + C K_z^{-\frac{1}{2}}) + C \left\| [\sum_{i=1}^{K_x} \sum_{j=1}^{K_y} \left[(\tilde{h}_x^K)^{m-1} \right]_{ij} \langle \tilde{\varphi}_j(x), f(.) \rangle \, \tilde{\varphi}_i(y)] + (\tilde{h}_x^{\perp})^{m-1} - (h_x)^{m-1}[.] \right\|_{op}
\end{aligned}
$$

where I used that the operator norm of $\tilde{h}_x^K$ and $\tilde{h}_x^{\perp}$ are bounded, so the operator norm of any finite product thereof is likewise bounded, by a bound C_{m-1} which may depend on m. Applying the above step inductively, obtain upper bound

$$C(\epsilon_K + C K_z^{-\frac{1}{2}}) \sum_{s=1}^{m-1} C_s$$

Furthermore, by the fact that, by construction, the spectral radius of matrix $\tilde{h}_x^K$ is less than one, and likewise for $\tilde{h}_x^{\perp}$. There exists some M such that for all $m > M$ $\left\| (\tilde{h}_x^K)^m \right\|_{op} < 1$, and this radius decays exponentially. Therefore, C_s, eventually decays geometrically, and so $\sum_{s=1}^{\infty} C_s < \infty$. So for any m, the upper bound is

$$O(\epsilon_K + K_z^{-\frac{1}{2}})$$

with constant that does not depend on K or m, as claimed.

The result for $g_x h_x^m$ follows analogously. □

Proof of Lemma (28)

Proof. First I show left multiplication by an invertible matrix leaves the solution unchanged, then show right multiplication acts only as a change of basis. The pencil (W^lB, W^lA) is equal to (W^lQ^*SU, W^lQ^*TU) by the generalized Schur decomposition of (B, A). Denoting $Q^lR^l = qr(W^lQ^*)$, where $R^l = \begin{bmatrix} R^l_{11} & R^l_{12} \\ 0 & R^l_{22} \end{bmatrix}$ conformably with the partitions of S and T, the generalized Schur decomposition of $(W^lB, W^lA) = (Q^lR^lSU, Q^lR^lTU)$. As a result, g_x derived from this decomposition, as it depends only on U is unchanged, and $h_x = (U_{11} + U_{12}g_x)^{-1}(R^l_{11}S_{11})^{-1}(R^l_{11}T_{11})(U_{11} + U_{12}g_x)$ which, by invertibility of R^l_{11}, is exactly h_x for the policy derived from (B, A). So, the policy derived from (W^lBW^r, W^lAW^r) is the same as that from (BW^r, AW^r). Denoting $Q^rR^r = qr(UW^l)$,

$$\begin{aligned}(BW^r, AW^r) &= (Q^*SUW^r, Q^*TUW^r) \\ &= (Q^*SR^rR^{r-1}UW^r, Q^*TR^rR^{r-1}UW^r) \\ &= (Q^*SR^rQ^r, Q^*TR^rQ^r)\end{aligned}$$

which is a valid generalized Schur decomposition, giving policy operators $g^r_x = -Q^{r-1}_{22}Q_{21}$, $h^r_x = (Q^r_{11} + Q^r_{12}g^r_x)^{-1}(S_{11}R^r_{11})^{-1}T_{11}R^r_{11}(Q^r_{11} + Q^r_{12}g^r_x)$. Replacing Q^r by $R^{r-1}UW^r$, obtain $g^r_x = -((R^{r-1})_{22}U_{22}W^r_y)^{-1}(R^{r-1})_{22}U_{21}W^r_x = W^{r-1}_y(-U^{-1}_{22}U_{21})W^r_x = W^{r-1}_y g_x W^r_x$ as claimed. Similarly

$$\begin{aligned}Q^r_{11} + Q^r_{12}g^r_x &= \begin{bmatrix} (R^{r-1})_{11} & (R^{r-1})_{22} \end{bmatrix} \begin{bmatrix} U_{11}W^r_x + U_{21}W^r_y g^r_x \\ U_{21}W^r_x + U_{22}W^r_y g^r_x \end{bmatrix} \\ &= R^{r-1}_{11}(U_{11} - U_{21}U^{-1}_{22}U_{21})W^r_x\end{aligned}$$

and so

$$\begin{aligned}h^r_x &= (R^{r-1}_{11}(U_{11} - U_{21}U^{-1}_{22}U_{21})W^r_x)^{-1}R^{r-1}_{11}S^{-1}_{11}T_{11}R^r_{11}R^{r-1}_{11}(U_{11} - U_{21}U^{-1}_{22}U_{21})W^r_x \\ &= W^{r-1}_x(U_{11} + U_{12}g_x)^{-1}S^{-1}_{11}T_{11}(U_{11} + U_{12}g_x)W^r_x = W^{r-1}_x h_x W^r_x\end{aligned}$$

as claimed. □