Impulse Response Matching Estimators for DSGE Models

Pablo Guerron-Quintana* Atsushi Inoue† Lutz Kilian‡

Federal Reserve Bank of Philadelphia | Vanderbilt University | University of Michigan CEPR

July 8, 2015

Abstract

One of the leading methods of estimating the structural parameters of DSGE models is the VAR-based impulse response matching estimator. The existing asympotic theory for this estimator does not cover situations in which the number of impulse response parameters exceeds the number of VAR model parameters. Situations in which this order condition is violated arise routinely in applied work. We establish the consistency of the impulse response matching estimator in this situation, we derive its asymptotic distribution, and we show how this distribution can be approximated by bootstrap methods. Our methods of inference remain asymptotically valid when the order condition is satisfied, regardless of whether the usual rank condition for the application of the delta method holds. Our analysis sheds new light on the choice of the weighting matrix and covers both weakly and strongly identified DSGE model parameters. We also show that under our assumptions special care is needed to ensure the asymptotic validity of Bayesian methods of inference. A simulation study suggests that the frequentist and Bayesian point and interval estimators we propose are reasonably accurate in finite samples. We also show that using these methods may affect the substantive conclusions in empirical work.

JEL Classification Number: C32, C52, E30, E50
Key Words: Structural estimation, DSGE, VAR, impulse response, nonstandard asymptotics, bootstrap, weak identification, robust inference.

*Federal Reserve Bank of Philadelphia, Philadelphia, PA 19106-1574. Email: pguerron@gmail.com.
†Department of Economics, Vanderbilt University, Nashville, TN 37235-1819. E-mail: atsushi.inoue@vanderbilt.edu
‡University of Michigan, Department of Economics, 238 Lorch Hall, Ann Arbor, MI 48109-1220. E-mail: lkilian@umich.edu.

1 Introduction

Structural impulse responses play a central role in modern macroeconomics. It is common to estimate the structural parameters of a dynamic stochastic general equilibrium (DSGE) models by choosing these parameters so as to minimize a suitably weighted average of the distance between the structural impulse responses implied by the DSGE model and the corresponding structural impulse responses implied by an approximating vector autoregressive (VAR) model fit to actual data. One advantage of this approach compared with full information maximum likelihood estimators of DSGE models is that it does not require the model to fit well in all dimensions, but allows the user to focus on the dimension of the model that matters most to macroeconomists (also see Dridi, Guay and Renault 2007; Hall et al. 2012).

Such impulse response matching estimators have been employed in Rotemberg and Woodford (1997), Altig, Christiano, Eichenbaum and Lindé (2011), Boivin and Giannoni (2006), Christiano, Eichenbaum and Evans (2005), DiCecio (2005), DiCecio and Nelson (2007), Dupor, Han and Tsai (2007), Iacoviello (2005), Jordà and Kozicki (2007) and Uribe and Yue (2006), among others. In related research, Christiano, Trabandt and Walentin (2011) propose a Bayesian version of the impulse response matching estimator in which the quasi-likelihood function based on the distance between VAR and DSGE model impulse responses is combined with prior information. Other applications of Bayesian impulse response matching estimators include Christiano, Eichenbaum and Trabandt (2013) and Kormilitsina and Nekipelov (2013).

Because impulse response matching estimators are classical minimum distance (CMD) estimators, by construction they inherit all potential problems associated with CMD estimation (see, e.g., Newey and Smith 2004). Notably, the use of the optimal weighting matrix induces finite-sample bias in the estimator, which is why most applied users rely on a diagonal weighting matrix instead. In this paper we identify another potential problem that is specific to impulse response matching estimators. In estimating the structural

parameters of DSGE models, macroeconomists often match response functions evaluated across many horizons such that the number of impulse response coefficients exceeds the dimensionality of the VAR model parameters (see, e.g., Iacoviello 2005; Uribe and Yue 2006; Altig, Christiano, Eichenbaum and Lindé 2011). We show that this practice may cause the joint distribution of the structural impulse responses to be singular, which in turn renders the asymptotic behavior of the resulting impulse response matching estimator nonstandard. As a result, standard asymptotic and finite-sample results for CMD estimators no longer apply. We develop an alternative asymptotic theory of the impulse response matching estimator for this practically relevant context. Our paper makes four distinct theoretical contributions.

First, we show that in this case the impulse response matching estimator has a nonstandard convergence rate when using the optimal weighting matrix. While the estimator remains consistent, its asymptotic distribution is nonstandard. Both the rate of convergence and the nonnormality of the asymptotic distribution differ from standard results for CMD estimators. We establish that the nonstandard asymptotic approximation may be recovered by bootstrap methods. Of course, in the absence of asymptotic normality, one would not want to report standard errors for this estimator, but rely on bootstrap confidence intervals that do not rely on asymptotic normality.

In contrast, the impulse response matching estimator based on the diagonal weighting matrix remains $\sqrt{T}$-consistent and asymptotically normal as in the standard CMD case. The asymptotic variance of the latter estimator, however, differs from the case in which the number of impulse responses to be matched is no larger than the number of VAR model parameters. We show that the asymptotic variance may nevertheless be approximated by the same bootstrap methods as in the case in which the dimensionality of the impulse response vector is no larger than that of the VAR model parameters. The latter result provides a formal justification for the use of the diagonal weighting matrix in applied work in a case not covered by existing asymptotic theory.

Second, our asymptotic results matter not only for the construction of point and interval estimates for structural parameters. We also prove that conventional tests of overidentifying restrictions, as employed in Boivin and Giannoni (2006), for example, have a nonstandard asymptotic distribution when the number of impulse response parameters exceeds the number of VAR model parameters, invalidating the use of conventional critical values.

Third, our work also has implications for the use of Bayesian impulse response matching estimators. Often in the literature, Bayesian estimators are used as a convenient device for constructing asymptotic approximations. It may be tempting to base inference on a point estimate constructed from the mean, median or mode of the quasi-posterior of the structural parameters together with an estimate of the asymptotic standard error based on the standard deviation of this distribution. Although Markov Chain Monte Carlo methods may indeed be used to construct point estimators of the structural parameters based on the mean, median or mode, we show that one cannot use the standard deviations of the quasi-posterior distribution to approximate the asymptotic standard errors of the structural parameter estimator when the number of impulse responses exceeds the number of VAR model parameters. This is true whether one employs the optimal weighting matrix or the diagonal weighting matrix. In contrast, the frequentist impulse response matching estimator based on the diagonal weighting matrix allows consistent estimation of the asymptotic standard errors by bootstrap methods, for example. Alternatively, asymptotically valid Bayesian inference may be conducted by constructing the variance using the sandwich formula of Chernozhukov and Hong (2003).

We observe that the latter two approaches remain asymptotically valid even when there are fewer impulse response parameters than VAR model parameters, regardless of whether the rank condition holds or not. This point is important because the order condition in question is only a necessary and not a sufficient condition for the validity of the conventional asymptotics of impulse response matching estimators. In practice, the rank condition for

the application of the delta method may fail, even when the order condition is satisfied. Using methods of inference on structural parameters that are invariant to the failure of the rank condition is important, given the difficulty of verifying this condition in practice, especially in larger-scale macroeconomic models.

Fourth, it is well known that structural parameters of macroeconomic models may not be strongly identified. This problem also afflicts impulse response matching estimators, as documented in Canova and Sala (2009). We propose a nonstandard confidence interval for the structural parameters of the underlying data generating process that is robust to weak identification problems.

The remainder of the paper is organized as follows. Section 2 examines the consistency and asymptotic distribution of the impulse response matching estimators in question and proposes suitable bootstrap methods of inference. Because both the impulse response matching estimator based on the optimal weighting matrix and the estimator based on the diagonal weighting matrix are practically feasible and asymptotically valid, the question arises which approach implies more accurate confidence intervals for structural model parameters in finite samples. Section 3 evaluates the quality of these asymptotic approximations based on a Monte Carlo simulation experiment. Based on a small-scale New Keynesian model we provide some tentative evidence that confidence intervals for the structural parameters based on the diagonal weighting matrix tend to be slightly more accurate overall than intervals based on the optimal weighting matrix. They also appear more robust to the choice of the VAR lag order and the maximum horizon of the impulse response function. Moreover, the use of a diagonal weighting matrix typically implies point estimates with lower mean-squared error (MSE) than alternative estimators, even allowing for corrections for any finite-sample VAR model misspecification.

In Section 4, we illustrate the implementation of the proposed methods in the context of a prototypical medium-scale New Keynesian DSGE model of the type used at many central banks. This empirical example illustrates that basing estimates of the asymptotic standard

error on the standard deviation of the quasi-posterior of the structural parameters tends to understate the uncertainty in the structural parameter estimates. For example, whereas the point estimate of the price-markup factor is quite robust to the choice of method, its standard error is about three times as large one would have concluded based on the standard deviation of the quasi-posterior. These results are based on the conventional premise in empirical work that the structural parameters are strongly identified. We also present alternative estimates that take account of the possibility that some parameters are only weakly identified. We illustrate that allowing for weak identification in some cases affects the substantive conclusions, while in others it does not. The concluding remarks are in Section 5. The proofs are contained in the appendix.

2 Asymptotic Theory

The thought experiment is that the data are generated by a DSGE model. At least some of the structural parameters of this DSGE model are unknown. The DSGE model is approximated by a finite-order structural VAR model with identifying restrictions that are consistent with the underlying DSGE model.[1] The objective is to recover an estimate of the unknown structural parameters in the DSGE model by searching the space of these parameters for the parameter values that result in the closest match between the structural VAR impulse responses based on the actual data and those from the DSGE model evaluated at the hypothesized parameter values. We are concerned with the asymptotic properties of this impulse response matching estimator in repeated sampling. As is standard in this literature, it is assumed that the structural impulse responses obtained from the VAR model are strongly identified.

Let γ_0 denote the $q \times 1$ vector of the population structural impulse responses (excluding all impulse responses that are not estimated). Then the structural impulse response

[1]Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson (2007) make precise the conditions under which a DSGE model may be approximated by a finite-order VAR model.

estimator $\widehat{\gamma}_T$ of γ_0 is obtained from an estimated VAR model fitted to the actual data, its bootstrap analogue $\widehat{\gamma}_T^*$ is obtained from a VAR model fitted to data simulated from the estimated VAR model, and the double-bootstrap estimator $\widehat{\gamma}_T^{**}$ is obtained from a VAR model fitted to data simulated from the VAR model evaluated at the bootstrap parameter estimates. Closed-form solutions for the structural impulse response estimator from VAR models are provided in Lütkepohl (1990). Because structural impulse responses are functions of the slope parameters and of the error covariance matrix of the VAR model, which in turn are an implicit function of the first and second moments of the data, we can write $\gamma_0 = \gamma(\mu)$, $\widehat{\gamma}_T = \gamma(\bar{X}_T)$, $\widehat{\gamma}_T^* = \gamma(\bar{X}_T^*)$ and $\widehat{\gamma}_T^{**} = \gamma(\bar{X}_T^{**})$ where μ, $\bar{X}_T$, $\bar{X}_T^*$ and $\bar{X}_T^{**}$ are $k \times 1$ vectors of the population moments, the sample moments, the bootstrap sample moments and the double-bootstrap sample moments, respectively. k corresponds to the number of VAR model parameters, defined as the total number of slope parameters plus the number of unique elements in the error covariance matrix.

In impulse response matching estimation, an $l \times 1$ vector of structural parameters of a macroeconomic model, θ, is estimated based on a restriction of the form:

$$\gamma_0 \;=\; f(\theta_0). \tag{1}$$

In this paper, we consider two types of impulse response matching estimators. One is based on the optimal weighting matrix,

$$\widehat{\theta}_{opt,T} \;=\; \text{argmin}_{\theta\in\Theta}(\widehat{\gamma}_T - f(\theta))'\widehat{\Sigma}_T^{*-1}(\widehat{\gamma}_T - f(\theta)), \tag{2}$$

where $\widehat{\Sigma}_T^*$ is the bootstrap covariance matrix estimator of impulse responses,

$$\widehat{\Sigma}_T^* \;=\; \frac{1}{B}\sum_{b=1}^{B}(\widehat{\gamma}_T^{*(b)} - \bar{\widehat{\gamma}}_T^*)(\widehat{\gamma}_T^{*(b)} - \bar{\widehat{\gamma}}_T^*)', \tag{3}$$

$\widehat{\gamma}_T^{*(b)}$ is the bth bootstrap estimator of impulse responses for $b = 1, 2, ..., B$, and $\bar{\widehat{\gamma}}_T^* =$

$(1/B)\sum_{b=1}^{B}\widehat{\gamma}^{*(b)}$. The other is based on the suboptimal weighting matrix

$$\widehat{\theta}_{diag,T} \;=\; \text{argmin}_{\theta\in\Theta}(\widehat{\gamma}_T - f(\theta))'W_T(\widehat{\gamma}_T - f(\theta)), \tag{4}$$

where W_T is a positive definite matrix. The leading example of W_T is the diagonal matrix whose diagonal elements are the reciprocals of the diagonal elements of $\widehat{\Sigma}_T^*$.

2.1 Asymptotic Properties of Impulse Response Matching Estimators of Structural Model Parameters

We initially focus on a situation in which the structural parameters are strongly identified. Our results are based on stochastic expansions and matrix decompositions. Suppose that there are conformable matrices B_0,...,B_H such that

$$\begin{aligned} T^{\frac{1}{2}}(\gamma(\bar{X}_T) - \gamma(\mu)) \;&=\; B_0 Z_T + T^{-\frac{1}{2}}B_1(Z_T\otimes Z_T) \\ &\quad + \cdots + T^{-\frac{H}{2}}B_H(Z_T\otimes\cdots\otimes Z_T) + o_p(T^{-\frac{H}{2}}), \end{aligned} \tag{5}$$

where H is the maximum horizon of the impulse responses to be matched. Conventional delta-method asymptotics are based on the leading term of this expansion. If $q > k$, the first-order asymptotic variance, B_0B_0', will be singular. It follows from the Schur decomposition theorem (Theorem 13 of Magnus and Neudecker, 1999, p.16) that there exists an orthonormal matrix S whose columns are eigenvectors of B_0B_0' and a diagonal matrix Λ whose diagonal elements are the eigenvalues of B_0B_0' such that

$$S'B_0B_0'S \;=\; \Lambda. \tag{6}$$

Stack the eigenvectors associated with the k largest eigenvalues of B_0B_0' in S_0 and let $q_0 = k$. Using a subset of the $q - q_0$ remaining eigenvectors that are not used in S_0, form S_1 so that $S_1'B_1$ contains no row vector of zeros. Let q_1 denote the number of columns in

S_1. Using a subset of the $q - q_0 - \cdots - q_{j-1}$ remaining eigenvectors that are not used in S_0, S_1,...S_{j-1}, form S_j so that $S_j' B_j$ contains no row vector of zeros. Let q_j be the number of columns in S_j. Stop when $q_0 + q_1 + \cdots + q_r = q$. Then $S = [S_1\ S_2\ \cdots\ S_r]$. The stochastic expansion (5) and the decomposition (6) allow us to analyze the asymptotic behavior of the impulse response matching estimator when the number of impulse response parameters exceeds the number of VAR model parameters.

To be more precise, our analysis requires the following conditions.

Assumptions.

(a) $Z_T \equiv \Omega^{-\frac{1}{2}}\sqrt{T}(\bar{X}_T - \mu) \xrightarrow{d} Z \equiv \mathcal{N}(0, I_k)$ and $Z_T^* \equiv \Omega^{-\frac{1}{2}}\sqrt{T}(\bar{X}_T^* - \bar{X}_T) \xrightarrow{d} Z^* \equiv \mathcal{N}(0, I_k)$ where Ω is a positive definite matrix and the convergence of Z_T^* is with respect to the bootstrap probability measure conditional on the data.

(b) There are conformable matrices B_0,...,B_H and $\widehat{B}_0$,...,$\widehat{B}_H$ such that

$$\begin{aligned} T^{\frac{1}{2}}(\gamma(\bar{X}_T) - \gamma(\mu)) &= B_0 Z_T + T^{-\frac{1}{2}} B_1 (Z_T \otimes Z_T) \\ &\quad + \cdots + T^{-\frac{H}{2}} B_H (Z_T \otimes \cdots \otimes Z_T) + o_p(T^{-\frac{H}{2}}), \qquad (7) \\ T^{\frac{1}{2}}(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T)) &= \widehat{B}_0 Z_T^* + T^{-\frac{1}{2}} \widehat{B}_1 (Z_T^* \otimes Z_T^*) \\ &\quad + \cdots + T^{-\frac{H}{2}} \widehat{B}_H (Z_T^* \otimes \cdots \otimes Z_T^*) + o_p^*(T^{-\frac{H}{2}}), \qquad (8) \end{aligned}$$

where $\widehat{B}_j = B_j + o_p(1)$ for $j = 0, 1, ..., H$, and $o_p^*(T^{-\frac{H}{2}})$ is defined with respect to the bootstrap probability measure conditional on the data with probability one.

(c)

$$\Upsilon_T S'(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T))(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T))' S \Upsilon_T$$

is uniformly integrable, where S is the orthonormal matrix whose columns are eigenvectors of $B_0 B_0'$ such that $S' B_0 B_0' S = \Lambda$ and Λ is a diagonal matrix whose diagonal

elements are the eigenvalues of $B_0 B_0'$, and where

$$\Upsilon_T = \begin{bmatrix} T^{\frac{1}{2}} I_{q_0} & 0_{q_0 \times q_1} & \cdots & 0_{q_0 \times q_r} \\ 0_{q_1 \times q_0} & T I_{q_1} & \cdots & 0_{q_1 \times q_r} \\ \vdots & \vdots & \ddots & \vdots \\ 0_{q_r \times q_0} & 0_{q_r \times q_1} & \cdots & T^{\frac{r+1}{2}} I_{q_r} \end{bmatrix}.$$

(d) $J = E(\xi\xi')$ is nonsingular where

$$\xi = \begin{bmatrix} S_0' B_0 Z \\ S_1' B_1 (Z \otimes Z) \\ \vdots \\ S_r' B_r (Z \otimes \cdots \otimes Z) \end{bmatrix},$$

and $S_1, ..., S_r$ are submatrices of S such that $S = [S_0 \; S_1 \; \cdots \; S_r]$.

(e) There is a unique value of θ, θ_0, in the interior of a compact set in $\Re^l$, Θ, such that $\gamma(\mu) = f(\theta_0)$, where $f : \Theta \to \Re^q$ is continuously differentiable in θ in the interior of Θ. The Jacobian matrix of f at θ_0, $F_0 = F(\theta_0)$, where $F(\theta) = \partial f(\theta)/\partial\theta'$, has rank l.

(f) W_T and W_T^* are sequences of positive definite matrices such that $W_T = W + o_p(1)$ and $W_T^* = W_T + o_p^*(1)$ where W is positive definite.

(g) (1) For the optimal weighting matrix, if $q > k$, then θ_0 is the unique value of θ in Θ that satisfies $S_r'(f(\theta_0) - f(\theta)) = 0_{q_r \times 1}$ where S_r is the $q \times q_r$ submatrix of S that consists of the last q_r columns of S as defined in the proof of Theorem 1.

(2) For the diagonal weighting matrix, $F_0' W B_0$ has rank l.

Remarks. 1. Assumption (a) holds, for example, when applying the residual-based bootstrap to stationary homoskedastic vector autoregressive processes.

2. Assumption (b) follows from a Taylor series expansion of the left-hand side of equations (7) and (8). The delta method is based on the first-order term of the stochastic expansion on the right-hand side. The higher-order stochastic terms on the right-hand side have also been used to develop Edgeworth expansions of the distribution of estimators (see Hall 1992). Assumption (b) holds, for example, for stationary vector autoregressive processes with positive definite error covariance matrices and short-run exclusion restrictions. For more primitive assumptions for the existence of asymptotic expansions of the distribution of estimators in stationary time series models see Bao and Ullah (2007) and Bao (2007).

3. Assumption (c) ensures that the bootstrap method can be used to estimate the limiting covariance matrix. The existence of S is guaranteed by the Schur decomposition theorem (Theorem 13 of Magnus and Neudecker, 1999, p.16).

4. Assumption (d) ensures that the impulse response matching estimator based on the optimal weighting matrix has a nonsingular asymptotic covariance matrix when scaled and rotated properly.

5. Assumptions (e) and (f) are standard assumptions for classical minimum distance estimators.

6. Assumption (g)(1) is for the impulse response matching estimator based on the optimal weighting matrix and is stronger than the identification condition in Assumption (e). Because $q_r < q$ can be smaller than l, there may be another value of θ at which $\gamma(\mu) - f(\theta) = f(\theta_0) - f(\theta) \neq 0_{q\times 1}$ and $S_r'(f(\theta_0) - f(\theta)) = 0_{q_r\times 1}$.

7. Assumption (g)(2) is for the impulse response matching estimator based on the diagonal weighting matrix. When $\text{rank}(B_0) = q$, it simplifies to the standard assumption that the Jacobian has full rank, i.e., $\text{rank}(F(\theta_0)) = l$.

Theorem 1 (Consistency of Impulse Response Matching Estimators). Suppose that Assumptions (a)–(g) hold. Then

(a) $\widehat{\theta}_{opt,T} \xrightarrow{p} \theta_0$.

(b) $\widehat{\theta}_{diag,T} \xrightarrow{p} \theta_0$.

Next, we derive the asymptotic distributions of the impulse response matching estimator.

Theorem 2 (Asymptotic Distributions of Impulse Response Matching Estimators). Suppose that Assumptions (a)–(g) hold. Then

(a)

$$T^{\frac{r+1}{2}}(\widehat{\theta}_{opt,T} - \theta_0) \xrightarrow{d} (F_0' S_r J^{rr-1} S_r' F_0)^{-1} F_0' S_r J^{r-1} \xi, \tag{9}$$

$$T^{\frac{r+1}{2}}(\widehat{\theta}^*_{opt,T} - \widehat{\theta}_{opt,T}) \xrightarrow{d} (F_0' S_r J^{rr-1} S_r' F_0)^{-1} F_0' S_r J^{r-1} \xi^*, \tag{10}$$

where J^{rr-1} is the $q_r \times q_r$ lower-right submatrix of J^{-1} and J^{r-1} is the $q_r \times q$ lower submatrix of J^{-1}, and

(b)

$$T^{\frac{1}{2}}(\widehat{\theta}_{diag,T} - \theta_0) \xrightarrow{d} (F_0' W F_0)^{-1} F_0' W B_0 Z, \tag{11}$$

$$T^{\frac{1}{2}}(\widehat{\theta}^*_{diag,T} - \widehat{\theta}_{diag,T}) \xrightarrow{d} (F_0' W F_0)^{-1} F_0' W B_0 Z^*. \tag{12}$$

Here the convergences in (10) and (12) are with respect to the bootstrap probability measure conditional on the data.

Remarks. Theorem 2(a) shows that the impulse response matching estimator based on the optimal weighting matrix has a nonstandard convergence rate that is faster than $T^{\frac{1}{2}}$ and has a nonstandard asymptotic distribution, when the number of impulse responses exceeds the number of VAR model parameters. While the bootstrap can mimic the convergence rate and nonstandard asymptotic distribution, the estimator has higher-order bias in that the limiting distribution has nonzero mean if r is even because of the properties of ξ^*. This fact makes it necessary to employ bootstrap confidence intervals that can accommodate

these features such as Hall's percentile interval. In contrast the impulse response matching estimator based on the diagonal weighting matrix in Theorem 2(b) is $\sqrt{T}$-consistent and asymptotically normal because Z and Z^* are normal.

2.2 Asymptotic Distributions of the Test Statistic for Overidentifying Restrictions

The results of section 2.1 not only have implications for the construction of point and interval estimates of the structural parameters, but they also affect tests of overidentifying restrictions. The conventional test statistic for overidentifying restrictions for testing the null hypothesis that $\gamma(\mu) = f(\theta_0)$ is defined as

$$J_T \;=\; (\widehat{\gamma}_T - f(\widehat{\theta}_{opt,T}))'\widehat{\Sigma}_T^{*-1}(\widehat{\gamma}_T - f(\widehat{\theta}_{opt,T})). \tag{13}$$

Under standard assumptions including $q \leq k$, the test statistic has an asymptotic χ^2 distribution under the null hypothesis. This test has been used, for example, in Boivin and Giannoni (2006). The bootstrap analogue of this test statistic is defined as

$$J_T^* \;=\; \left(\widehat{\gamma}_T^* - f(\widehat{\theta}_{opt,T}^*) - \widehat{\gamma}_T + f(\widehat{\theta}_{opt,T})\right)' \widehat{\Sigma}_T^{**-1} \left(\widehat{\gamma}_T^* - f(\widehat{\theta}_{opt,T}^*) - \widehat{\gamma}_T + f(\widehat{\theta}_{opt,T})\right), \tag{14}$$

where the term $\widehat{\gamma}_T - f(\widehat{\theta}_{opt,T})$ accomplishes the required recentering (see Hall and Horowitz 1996).

Theorem 3 (Asymptotic Distribution of the Test Statistic for Overidentifying Restrictions). Suppose that Assumptions (a)–(f) and (g)(1) hold. Then

$$J_T \;\xrightarrow{d}\; \eta'\eta, \tag{15}$$

$$J_T^* \;\xrightarrow{d}\; \eta^{*\prime}\eta^*, \tag{16}$$

where

$$\begin{aligned}\eta &= J^{-\frac{1}{2}}\xi - J^{-\frac{1}{2}}S_r'F_0(F_0'S_rJ^{rr-1}S_r'F_0)^{-1}F_0'S_rJ^{r-1}\xi,\\ \eta^* &= J^{-\frac{1}{2}}\xi^* - J^{-\frac{1}{2}}S_r'F_0(F_0'S_rJ^{rr-1}S_r'F_0)^{-1}F_0'S_rJ^{r-1}\xi^*.\end{aligned}$$

Theorem 3 shows that the asymptotic distribution of the test statistic for overidentifying restrictions is nonstandard if $q > k$, but can be mimicked by the bootstrap.

2.3 Asymptotic Distributions of Bayesian Impulse Response Matching Estimators

Next, we evaluate the Bayesian impulse response matching estimator of Christiano et al. (2011) and Christiano et al. (2013) from an asymptotic point of view. Christiano et al. motivate their approach as building on the analysis in Kim (2002) in particular. Define the quasi-posterior density as

$$p(\theta) = \frac{\exp(-q_T(\theta))\pi(\theta)}{\int_\Theta \exp(-q_T(\theta))\pi(\theta)d\theta} \tag{17}$$

where $\pi(\theta)$ is the prior density and

$$q_T(\theta) = \frac{1}{2}(\widehat{\gamma}_T - f(\theta))'\widehat{\Sigma}_T^{*-1}(\widehat{\gamma}_T - f(\theta)). \tag{18}$$

Under the standard assumptions the quasi-posterior density converges to the asymptotic distribution of the impulse response matching estimator. As in Chernozhukov and Hong (2003), the quasi-posterior density is concentrated in a $T^{-(r+1)/2}$ neighborhood around θ_0, which is characterized by the local parameter

$$h = T^{\frac{r+1}{2}}(\theta - \theta_0) + T^{\frac{r+1}{2}}(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0), \tag{19}$$

where ∇ denotes the gradient and ∇^2 the Hessian. $T^{\frac{r+1}{2}}$ is used because of the convergence rate of the impulse response matching estimator. Define the quasi-posterior density for h as

$$p_T^*(h) \;=\; T^{-\frac{l(r+1)}{2}} p_T\left(\frac{h}{T^{\frac{r+1}{2}}} + \theta_0 - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) \tag{20}$$

The next theorem establishes the asymptotic behavior of this quasi-posterior distribution.

Theorem 4 (Asymptotic Behavior of the Quasi-Posterior Distribution). Suppose that Assumptions (a)–(f) and (g)(1) hold. Then, using the notation of Chernozhykov and Hong (2003),

$$\|p_T^*(h) - p_\infty^*(h)\|_{TVM(\alpha)} \equiv \int_{H_T} (1 + \|h\|^\alpha)|p_T^*(h) - p_\infty^*(h)|dh \;=\; o_p(1), \tag{21}$$

where

$$\begin{aligned} H_T &= \{h \in \Re^l : h = T^{\frac{r+1}{2}}(\theta - \theta_0) - T^{\frac{r+1}{2}}(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0) \text{ for some } \theta \in \Theta\}, \\ p_\infty^*(h) &= \sqrt{\frac{|T^{-(r+1)}\nabla^2 q_T(\theta_0)|}{(2\pi)^l}} \exp\left(-\frac{1}{2T^{r+1}} h'\nabla^2 q_T(\theta_0) h\right). \end{aligned}$$

Remarks. 1. Theorem 4 shows that the quasi-posterior is concentrated around the impulse response matching estimator $\widehat{\theta}_T$ and will be locally asymptotically normal. Heuristically, this follows from a quadratic expansion of $q_T(\theta)$ around $\widehat{\theta}_T$ in a neighborhood of $\widehat{\theta}_T$:

$$q_T(\theta) \;\approx\; q_T(\widehat{\theta}_T) + \frac{1}{2}(\theta - \widehat{\theta}_T)'\nabla^2 q_T(\widehat{\theta}_T)(\theta - \widehat{\theta}_T), \tag{22}$$

where the linear term is zero due to the first-order condition for $\widehat{\theta}_T$. Because $q_T(\widehat{\theta}_T)$ is constant in θ, the quasi-posterior is approximately proportional to

$$\exp\left(-\frac{1}{2}(\theta - \widehat{\theta}_T)'\nabla^2 q_T(\widehat{\theta}_T)(\theta - \widehat{\theta}_T)\right), \tag{23}$$

provided $\pi(\theta_0) > 0$, which basically implies a normal density with mean $\widehat{\theta}_T$ and covariance matrix $[\nabla^2 q_T(\widehat{\theta}_T)]^{-1}$.

2. This result means that the posterior distribution in equation (23) is different from the asymptotic distribution of the impulse response matching estimator in equation (11). Hence, Markov Chain Monte Carlo draws, which are designed to characterize the quasi-posterior distribution, cannot be used to estimate asymptotic standard errors and confidence intervals when $q > k$. Under Assumption (g)(1), a necessary condition for estimating the asymptotic standard errors from the quasi-posterior is that $q \leq k$.

3. It can be shown that the same problem arises when Assumption (g)(1) is replaced by Assumption (g)(2). In that case, the quasi-posterior again is normal, but different from equation (11), so we cannot rely on the standard errors of the quasi-posterior for asymptotic inference. Instead, one has to evaluate the sandwich formula in Chernozhukov and Hong (2003, p. 307), whether $q > k$ or $q \leq k$. Likewise, it is not asymptotically valid to compute confidence bands based on the upper and lower percentiles of the quasi-posterior of the structural parameters.

2.4 Robust Inference in the Presence of Rank Deficiencies

Although both impulse matching estimators based on the optimal weighting matrix and impulse response matching estimators based on the diagonal weighting matrix allow asymptotically valid inference, there is a reason to favor the impulse response matching estimator based on the diagonal weighting matrix. The condition that $q \leq k$, on which the conventional asymptotic analysis of the impulse response matching estimator is based, is an order condition, which is only necessary, but not sufficient for the conventional asymptotics to apply. There are cases in which our alternative asymptotic analysis is required even when $q \leq k$ because the rank condition fails. For example, consider the bivariate reduced-form VAR(1) model

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} y_{1,t-1} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} e_{1t} \\ e_{2t} \end{bmatrix}. \tag{24}$$

Suppose that one is interested in matching the two-step-ahead impulse responses only. Without loss of generality suppose that the covariance matrix of $[e_{1t}\ e_{2t}]'$ is known to be the identity matrix, but that the slope parameters are unknown. Then the two-step-ahead structural impulse responses are

$$\begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} b_{11}^2 + b_{12}b_{21} & b_{11}b_{12} + b_{12}b_{22} \\ b_{21}b_{11} + b_{22}b_{21} & b_{21}b_{12} + b_{22}^2 \end{bmatrix}. \tag{25}$$

Because there are four impulse response parameters ($q = 4$) and four unknown VAR model parameters ($k = 4$), the order condition $q \leq k$ for a nonsingular joint distribution is satisfied. For the joint distribution of the structural responses to be nonsingular, the Jacobian has to be of full rank. The Jacobian matrix takes the form

$$\begin{bmatrix} 2b_{11} & b_{21} & b_{12} & 0 \\ b_{21} & 0 & b_{11} + b_{22} & b_{21} \\ b_{21} & b_{11} + b_{22} & 0 & b_{12} \\ 0 & b_{21} & b_{12} & 2b_{11} \end{bmatrix}. \tag{26}$$

This matrix, however, may be rank deficient for certain parameter values even when the order condition is satisfied. For example, it will have rank 2 when $b_{11} = b_{22} = 0$ and $b_{12} = b_{21} = \kappa \neq 0$, indicating that the rank condition is violated and that it is not possible to rely on conventional asymptotics. A similar point is made in Benkwitz et al. (2000). While this example is tractable, it is somewhat artificial. It nevertheless illustrates that in general it is not enough to check the order condition before applying conventional asymptotic analysis to impulse response matching estimators.

Verifying the rank condition in more complicated models is often difficult as well as tedious. Our analysis in section 2.2 implies that evaluating this rank condition is not required in practice because, as long as one uses the diagonal weighting matrix, the same bootstrap methods may be used to compute the asymptotic standard errors for the struc-

tural parameters, whether the rank condition for the application of the delta method holds or not. A similar robustness result is obtained for the Bayesian impulse response estimator in section 2.4, provided the sandwich formula of Chernozhukov and Hong (2003) is used.

2.5 Inference When Identification is Not Strong

A common problem in applied work is that some parameters of the DSGE model may not be strongly identified. In the GMM context, this problem was first discussed in Stock and Wright (2000). While several methods of inference have been developed that are robust to weak identification problems in DSGE models, none of these methods are designed for impulse response matching estimators (e.g., Guerron-Quintana, Inoue and Kilian 2013; Dufour, Khalaf and Kichian 2013; Qu 2014; Andrews and Mikusheva 2015).

Below we derive the asymptotic distribution of the Wald test statistic without assuming the identifiability of θ_0. Our results apply whether $q > k$ or $q \leq k$.

Proposition (Asymptotic Distributions of the Wald Test Statistic of the Structural Impulse Responses).

Suppose that Assumptions (a)–(d) hold. Under $H_0 : \gamma(\mu) = f(\theta_0)$ for some $\theta_0 \in \Theta$,

$$\mathcal{W}_T \xrightarrow{d} \xi' J^{-1} \xi, \tag{27}$$

$$\mathcal{W}_T^* \xrightarrow{d} \xi^{*\prime} J^{-1} \xi^* \tag{28}$$

where

$$\begin{aligned}
\mathcal{W}_T &= (\widehat{\gamma}_T - f(\theta_0)' \Sigma_T^{*-1} (\widehat{\gamma}_T - f(\theta_0)), \\
\mathcal{W}_T^* &= (\widehat{\gamma}_T^* - \widehat{\gamma}_T)' \Sigma_T^{**-1} (\widehat{\gamma}_T^* - \widehat{\gamma}_T), \\
\Sigma_T^* &= \frac{1}{B} \sum_{j=1}^{B} \left(\widehat{\gamma}_T^{*(j)} - \bar{\widehat{\gamma}}_T^* \right) \left(\widehat{\gamma}_T^{*(j)} - \bar{\widehat{\gamma}}_T^* \right)', \\
\Sigma_T^{**(j)} &= \frac{1}{B} \sum_{k=1}^{B} \left(\widehat{\gamma}_T^{**(j,k)} - \bar{\widehat{\gamma}}_T^{(j)**} \right) \left(\widehat{\gamma}_T^{**(j,k)} - \bar{\widehat{\gamma}}_T^{(j)**} \right)',
\end{aligned}$$

with $\widehat{\gamma}_T^{**(j,k)} = \gamma(\bar{X}_T^{**(j,k)})$ denoting the kth bootstrap draw of the structural impulse response based on the jth bootstrap estimate and $\bar{\widehat{\gamma}}_T^{**(j)} = (1/B)\sum_{k=1}^{B} \widehat{\gamma}_T^{**(j)}$. The convergence of $\mathcal{W}_T^{\gamma*}$ is with respect to the bootstrap probability measure conditional on the data with probability one.

This proposition follows from Theorem 2 in Inoue and Kilian (2014) by replacing $\gamma_0 = \gamma(\mu)$ with $f(\theta_0)$. Because the asymptotic distribution does not depend on the strength of identification of θ_0 and can be approximated by the bootstrap, one can invert the Wald statistic to obtain a $100(1-\varsigma)\%$ asymptotic confidence set for θ_0:

$$\{\theta \in \Theta : \ \mathcal{W}_T(\theta) \leq \mathcal{W}_{1-\varsigma}^*\} \tag{29}$$

where $\mathcal{W}_{1-\varsigma}^*$ is the $100(1-\varsigma)$ percentile of the bootstrap distribution of $\mathcal{W}_T^*(\theta)$. Pointwise confidence intervals for the individual elements of the structural parameter vector θ may be obtained by the projection method (see, e.g., Dufour and Taamouti 2005, Chaudhuri and Zivot 2011; Guerron-Quintana, Inoue and Kilian 2013).

3 Monte Carlo Simulation Experiments

For this section, we focus on a small-scale New Keynesian model that often serves as an illustrative example in the literature. This model consists of a Phillips curve, a Taylor rule, an investment-savings relationship, and the exogenous driving processes z_t and ξ_t:

$$\pi_t = \kappa x_t + \beta \mathbb{E}(\pi_{t+1}|\mathcal{I}_{t-1}), \tag{30}$$

$$R_t = \rho_r R_{t-1} + (1-\rho_r)\phi_\pi \pi_t + (1-\rho_r)\phi_x x_t + \xi_t, \tag{31}$$

$$x_t = \mathbb{E}(x_{t+1}|\mathcal{I}_{t-1}) - \sigma\left(\mathbb{E}(R_t|\mathcal{I}_{t-1}) - \mathbb{E}(\pi_{t+1}|\mathcal{I}_{t-1}) - z_t\right), \tag{32}$$

$$z_t = \rho_z z_{t-1} + \sigma^z \varepsilon_t^z, \tag{33}$$

$$\xi_t = \sigma^r \varepsilon_t^r, \tag{34}$$

where x_t, π_t and R_t denote the output gap, inflation rate, and interest rate, respectively. Note that this model has more variables than shocks. The structural shocks ε_t^z and ε_t^r are assumed to be distributed $\mathcal{NID}(0,1)$. The model parameters are the discount factor β, the intertemporal elasticity of substitution $1/\sigma$, the probability α of not adjusting prices for a given firm, the elasticity of substitution across varieties of goods, θ, the parameter ω controlling disutility of labor supply; ϕ_π and ϕ_x capture the central bank's reaction to changes in inflation and the output gap, respectively, and $\kappa = \frac{(1-\alpha)(1-\alpha\beta)}{\alpha}\frac{\omega+\sigma}{\sigma(\omega+\theta)}$.

While this model is similar to the model used in Guerron-Quintana et al. (2013), there is one crucial difference. In this model, inflation and real output do not react contemporaneously to the monetary policy shock, ξ_t, but they do respond contemporaneously to a shock to the investment-savings relationship, z_t. These restrictions are required for us to be able to identify the structural shocks of interest in the VAR model based on short-run identifying restrictions. Given this informational constraint, household and firms form expectations based on the information set $\mathcal{I}_{t-1}$.

Given the computational cost of evaluating higher-dimensional models, we focus on the estimation of one parameter only in the simulation study. For expository purposes, we concentrate on the problem of estimating the probability of not adjusting prices, α, by matching the impulse responses of inflation and of the interest rate with the remaining parameters set to their population values in estimation. The population parameters in the data generating process are $\sigma = 1$, $\alpha = 0.75$, $\beta = 0.99$, $\phi_\pi = 1.5$, $\phi_x = 0.125$, $\omega = 1$, $\rho_r = 0.75$, $\rho_z = 0.90$, $\theta = 6$, $\sigma^z = 0.30$, $\sigma^r = 0.20$. It can be shown that the parameter α is strongly identified.

This model may equivalently be expressed in the state-space representation

$$x_t = Ax_{t-1} + B\varepsilon_t, \tag{35}$$

$$y_t = Cx_t, \tag{36}$$

where x_t is a vector of state variables, ε_t is a vector that consists of the technology shock

and the monetary policy shock, and y_t is a vector that consists of inflation and the interest rate. A, B and C are matrices of suitable dimensions. Substitution of x_t in equation (36) yields the structural moving average representation of y_t in terms of current and lagged structural shocks ε_t. This moving average representation is invertible because, given our population parameter values, the eigenvalues of A are strictly less than unity in modulus, so Fernandez-Villaverde, Rubio-Ramirez, Sargent and Watson's (2007) condition for the existence of an infinite-order VAR representation is satisfied. This structural VAR(∞) model in turn may be approximated by a finite-order structural VAR model (see Inoue and Kilian 2002). Because the structural impact multiplier matrix of the VAR model, CB, is lower triangular given the informational constraints discussed earlier, we can recover the two structural shocks ε_t^z and ε_t^r by applying a lower triangular Cholesky decomposition to the residual covariance matrix with the diagonals of the decomposition normalized to be positive.

The Monte Carlo study consists of the following steps:

1. First, we generate 500 synthetic data sets of length T for inflation and the interest rate from the New Keynesian model evaluated at the true parameter values. We consider two alternative sample sizes: $T = 100$ and $T = 232$. The shorter sample corresponds to the length of a quarterly time series starting in 1984 with the onset of the Great Moderation. The longer sample corresponds to the period from 1950 to 2008, which represents another common situation in the empirical literature.

2. For each synthetic data set, we fit a bivariate VAR(p) model for inflation and the interest rate and estimate the four structural impulse response functions at horizons $0, ..., h-1$, which are stacked into a vector and denoted by $\widehat{\gamma}_T$. To this end, we use a Cholesky decomposition in which inflation does not react contemporaneously to the second structural shock in the VAR, which identifies this shock as the monetary policy shock. Using the standard nonparametric residual-based bootstrap for VAR models, we bootstrap the VAR(p) model and estimate a vector of bootstrap structural impulse

responses $\widehat{\gamma}_T^{*(j)}$ for $j = 1, ..., B$, where $B = 500$. For each of the B bootstrapped VAR parameter estimates, $\widehat{\gamma}_T^{*(j)}$, the residual-based bootstrap is applied and B vectors of bootstrap structural impulse responses, $\widehat{\gamma}_T^{**(j,b)}$, are computed for $b = 1, 2, ..., B$.

3. We estimate α. The other parameters are treated as known in the estimation for computational simplicity. Define $\widehat{\alpha}_T$ by

$$\widehat{\alpha}_{opt,T} = \text{argmin}_{\alpha \in A}(\widehat{\gamma}_T - g(\alpha))' W_{opt,T}(\widehat{\gamma}_T - g(\alpha)), \quad (37)$$

$$\widehat{\alpha}_{diag,T} = \text{argmin}_{\alpha \in A}(\widehat{\gamma}_T - g(\alpha))' W_{diag,T}(\widehat{\gamma}_T - g(\alpha)), \quad (38)$$

where $g(\alpha)$ is a vector of structural impulse responses implied by the macroeconomic model evaluated at α, $W_{opt,T} = [(1/B)\sum_{j=1}^{B}(\widehat{\gamma}_T^{*(j)} - \bar{\widehat{\gamma}}_T)(\widehat{\gamma}_T^{*(j)} - \bar{\widehat{\gamma}}_T^{*})']^{-1}$ and $W_{diag,T}$ is the diagonal matrix whose diagonal elements are given by the reciprocal of the diagonal elements of $(1/B)\sum_{j=1}^{B}(\widehat{\gamma}_T^{*(j)} - \bar{\widehat{\gamma}}_T)(\widehat{\gamma}_T^{*(j)} - \bar{\widehat{\gamma}}_T^{*})'$.

4. Let $\widehat{\alpha}_T^*$ be the bootstrap analogue of $\widehat{\alpha}_T$. Then

$$\widehat{\alpha}_{opt,T}^{*(j)} = \text{argmin}_{\alpha \in A}(\widehat{\gamma}_T^{*(j)} - g(\alpha) - \widehat{\gamma}_T + g(\widehat{\alpha}_{opt,T}))' W_{opt,T}^{*(j)}(\widehat{\gamma}_T^{*(j)} - g(\alpha) - \widehat{\gamma}_T + g(\widehat{\alpha}_{opt,T})), \quad (39)$$

$$\widehat{\alpha}_{diag,T}^{*(j)} = \text{argmin}_{\alpha \in A}(\widehat{\gamma}_T^{*(j)} - g(\alpha) - \widehat{\gamma}_T + g(\widehat{\alpha}_{diag,T}))' W_{diag,T}^{*(j)}(\widehat{\gamma}_T^{*(j)} - g(\alpha) - \widehat{\gamma}_T + g(\widehat{\alpha}_{diag,T})), \quad (40)$$

where $W_{opt,T}^{*(j)} = [(1/B)\sum_{k=1}^{B}(\widehat{\gamma}_T^{**(j,k)} - \bar{\widehat{\gamma}}_T^{*(j)})(\widehat{\gamma}_T^{**(j,k)} - \bar{\widehat{\gamma}}_T^{*(j)})']^{-1}$ and $W_{diag,T}^{*(j)}$ is the diagonal matrix whose diagonal elements are given by the reciprocal of the diagonal elements of $(1/B)\sum_{j=1}^{B}(\widehat{\gamma}_T^{**(j,k)} - \bar{\widehat{\gamma}}_T^{*(j)})(\widehat{\gamma}_T^{**(j,k)} - \bar{\widehat{\gamma}}_T^{*(j)})'$. Using these bootstrap estimates we construct nominal 68% and 90% confidence intervals based on Hall's (1992) percentile interval to allow for the fact that the asymptotic distribution may not be normal, depending on the choice of weighting matrix.

Our simulation evidence is necessarily tentative, but nevertheless provides some useful insights. Table 1 summarizes the properties of the point estimator of α. Because the quality

of the VAR approximation depends on the lag order p, we report results for a range of p. We also consider a range of values for H to allow for situations in which the asymptotic theory developed on this paper applies as well as for situations in which the conventional asymptotic theory for impulse response matching estimators applies.

The upper panel of Table 1 shows that the estimator based on the diagonal weighting matrix tends to have a small negative bias, whereas that based on the optimal weighting matrix tends to have somewhat larger positive bias, consistent with the presence of a higher-order bias in the theoretical analysis of section 2.1. The estimator based on the optimal weighting matrix also tends to have higher MSE than the estimator based on the diagonal weighting matrix.

These results are based on the value of $\gamma(\alpha)$ implied by the underlying macroeconomic model, which has a VAR(∞) representation. This creates a mismatch with the structural impulse responses estimated based on a finite-order approximation to this VAR(∞) model. We can refine the analysis by deriving in population the value of $\gamma(\alpha)$ based on the finite-order VAR representation of the macroeconomic model with the same choice of p as in the empirical VAR model. Let $\Gamma_j = E(y_t y_{t-j}')$ denote the population autocovariances implied by the state space representation given a structural parameter value. Then the population parameter values of the VAR(p) model fitted to data generated by the model may be

expressed as:

$$\underset{2p\times 2}{\Phi} = \begin{bmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & \cdots & \Gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p-1}' & \Gamma_{p-2}' & \cdots & \Gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_1' \\ \Gamma_2' \\ \vdots \\ \Gamma_p' \end{bmatrix}, \tag{41}$$

$$\underset{2\times 2}{\Sigma} = \Gamma_0 - \begin{bmatrix} \Gamma_1 & \Gamma_2 & \cdots & \Gamma_p \end{bmatrix} \begin{bmatrix} \Gamma_0 & \Gamma_1 & \cdots & \Gamma_{p-1} \\ \Gamma_1' & \Gamma_0 & \cdots & \Gamma_{p-2} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{p-1}' & \Gamma_{p-2}' & \cdots & \Gamma_0 \end{bmatrix}^{-1} \begin{bmatrix} \Gamma_1' \\ \Gamma_2' \\ \vdots \\ \Gamma_p' \end{bmatrix}. \tag{42}$$

The population structural impulse responses can be calculated from the slope coefficients Φ and the reduced-form error covariance matrix Σ.

The revised results are presented in the lower panel of Table 1. After correcting for VAR specification mismatch, the estimator based on the optimal weighting matrix becomes less sensitive to the choice of p and H. Nevertheless, overall, by the MSE metric the uncorrected estimator based on the diagonal weighting matrix is preferred. The finite-sample correction actually increases the bias, variance and MSE of the estimator of the structural parameter based on the diagonal weighting matrix. We conclude that overall the impulse response matching estimator based on the diagonal weighting matrix without corrections for any finite-sample VAR model misspecification is most reliable.

Table 2 reports the effective coverage probabilities of the corresponding nominal 68% and 90% bootstrap confidence intervals for α. The confidence intervals based on the diagonal weighting matrix tend to be more robust to the choice of p and H than the confidence intervals based on the optimal weighting matrix. The latter interval may perform erratically in some cases. Finite-sample corrections in those cases substantially improve the accuracy of the interval based on the optimal weighting matrix, but overall, the intervals implied by the estimator based on the diagonal weighting matrix tend to be even more ac-

curate with finite-sample corrections having an ambiguous effect on the coverage accuracy of the interval. Generally, the coverage rates are reasonably close to their nominal levels.

4 Empirical Application

For the empirical application, we consider a prototypical medium-scale New Keynesian DSGE model (see, e.g., Christiano, Eichenbaum, and Evans 2005; Smets and Wouters 2007; Altig, Christiano, Eichenbaum and Lindé 2011; Guerron-Quintana, Inoue, and Kilian 2013). Since this class of models has been extensively discussed in the macroeconomics literature, we provide only a brief summary. The main features of the model are as follows: The economy grows along a stochastic path; prices and wages are assumed to be sticky à la Calvo; preferences display internal habit formation; investment is costly; and finally, there are three sources of uncertainty: neutral and capital embodied technology shocks, and monetary shocks.

4.1 Households

The economy is populated by a continuum of households. Every period households must decide how much to consume, work, and invest. In addition, they must choose the amount of government bonds. Agents in the economy have access to complete markets; such an assumption is needed to eliminate wealth differentials arising from wage heterogeneity. Households maximize the expected present discounted value of utility

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left[\log(C_t - bC_{t-1}) - A \int_0^1 \frac{h_{j,t}^{1+\upsilon}}{1+\upsilon} dj \right] \tag{43}$$

subject to

$$P_t C_t + P_t \left(I_t + a(u_t)\overline{K}_t \right) / \Psi_t + B_{t+1} = R_t^K u_t \overline{K}_t + \int_0^1 W_{j,t} h_{j,t} dj + R_{t-1} B_t + \Pi_t + T_t,$$

and

$$\overline{K}_{t+1} = (1-\delta)\overline{K}_t + I_t\left(1 - S(\frac{I_t}{I_{t-1}})\right).$$

Here, $\mathbb{E}_t$ is the time t expectation operator conditional on the information set of the household; preferences display internal habit formation measured by $b \in (0,1)$; and $S(.)$ is a function reflecting the costs associated with adjusting the investment portfolio. This function is assumed to be increasing and convex satisfying $S = S = 0$ and $S'' > 0$ in the steady state. T_t corresponds to lump-sum transfers from the government to the household. B_t is the individual demand for one-period government bonds, which pay the gross nominal interest rate R_t. As in the related literature, it is assumed that physical capital can be used with different intensities (see, e.g., Christiano, Eichenbaum, and Evans; 2005). Using capital with intensity u_t yields the return $R_t^K u_t \overline{K}_t$ but entails a cost $a(u_t)$, which satisfies $a(1) = 0; a''(1) > 0; a'(1) > 0$. For future reference, we define $\sigma_a = a''(1)/a'(1)$. Finally, Π_t corresponds to profits from producers. Ψ_t is an investment-specific disturbance, which, following the literature, is assumed to grow at rate $\mu_{\Psi,t} = log(\Psi_t/\Psi_{t-1})$, where

$$\mu_{\Psi,t} = (1-\rho_\psi)\mu_\psi + \rho_{\mu,\psi}\mu_{\Psi,t-1} + \sigma_{\epsilon,\psi}\epsilon_{\Psi,t},$$

and $\epsilon_{\Psi,t}$ is distributed $\mathcal{NID}(0,1)$.

4.2 Wage Setting

Households sell differentiated labor services $h_{j,t}$ to a competitive firm that aggregates labor and sells it to final firms. This labor aggregator pays $W_{j,t}$ for each unit of differentiated labor of type j. The technology used by the aggregator is

$$H_t = \left[\int_0^1 h_{j,t}^{1/\lambda_w} dj\right]^{\lambda_w}, \quad 1 < \lambda_w.$$

It is straightforward to show that the relationship between the labor aggregate, H_t, and the aggregate wage, W_t, is given by

$$h_{j,t} = \left[\frac{W_{j,t}}{W_t}\right]^{-\lambda_w/(\lambda_w-1)} H_t.$$

To induce wage sluggishness, it is assumed that there exists a labor union representing all workers of type j. Each period, the union sets wages in a Calvo fashion. In particular, with exogenous probability ξ_w a union does not re-optimize wages each period. In that case, wages are set according to the rule of thumb $W_{j,t} = \pi^{1-\iota_w} (\pi_t)^{\iota_w} W_{j,t-1}\mu_{z^+}$. Here, μ_{z^+} is the average growth rate of the economy, as defined below, and ι_w is the degree of wage indexation to inflation.

4.3 Firms

There is a continuum of monopolistically competitive firms indexed by $i \in [0, 1]$, each producing an intermediate good from capital services, $k_{i,t}$, and labor services, $H_{i,t}$. Firms rent capital and labor in perfectly competitive factor markets. The production function is given by

$$Y_{i,t} = k_{i,t}^{\alpha} (z_t H_{i,t})^{1-\alpha} - z_t^{+}\psi,$$

where ψ is a fixed cost of production. The technology shock, z_t, grows at rate $\mu_{z,t} = \log(z_t/z_{t-1})$, which is assumed to follow the process

$$\mu_{z,t} = \mu_z + \sigma_{\epsilon,\mu}\epsilon_{z,t},$$

where $\epsilon_{z,t}$ is distributed $\mathcal{NID}(0, 1)$. The aggregate trend $z_t^{+} = \Psi_t^{\alpha/(1-\alpha)} z_t$ grows at rate $\mu_{z^+,t}$.[2]

Intermediate firms must borrow the wage bill in advance. As a consequence, the cost of

[2]The growth term in the fixed cost is needed for a well-defined steady state to exist about which the model can be solved.

hiring one unit of labor is $W_t R_t$. These firms choose prices to maximize the present value of profits; prices are set in Calvo fashion; i.e., each period, firms optimally revise their prices with an exogenous probability $1-\xi_p$. If, instead, a firm does not re-optimize its price, then the price is updated according to the rule: $P_{i,t} = \pi P_{i,t-1}$. Here, π is steady-state inflation.

There is a competitive firm that produces the final good using intermediate goods according to the technology

$$Y_t = \left[\int_0^1 Y_{j,t}^{1/\lambda_p} dj\right]^{\lambda_p}.$$

The parameter λ_p determines the degree of monopoly power enjoyed by intermediate producers.

4.4 Government

As in most of the recent New Keynesian literature, we assume a cashless economy (see Woodford 2003). The monetary authority sets the short-term interest rate according to a Taylor rule. In particular, the central bank smoothes interest rates and responds to deviations of actual inflation from steady-state inflation, π, and deviations of output from its target level, $\mathcal{Y}$.

$$\frac{R_t}{R} = \left(\frac{R_{t-1}}{R}\right)^{\rho_r} \left[\left(\frac{\pi_t}{\pi}\right)^{\phi_\pi} \left(\frac{\mathcal{Y}_t}{\mathcal{Y}}\right)^{\phi_y}\right]^{1-\rho_r} \exp(\sigma_{\epsilon,r}\epsilon_{r,t}). \tag{44}$$

The term $\epsilon_{r,t}$ is a random shock to the systematic component of monetary policy and is assumed to be standard normal; $\sigma_{\epsilon,r}$ is the standard deviation of the monetary shock. Following Christiano, Trabandt and Walentin (2011), $\mathcal{Y}_t$ corresponds to de-trended GDP such that $\mathcal{Y}_t = \frac{C_t + I_t/\Psi_t + G_t}{z_t^+}$. R is the steady-state gross nominal interest rate. Finally, we assume that government spending is given by $G_t = g z_t^+$. Here, g is a constant and the government uses lump-sum taxes to finance its purchases.

4.5 Estimation

We estimate the model in two stages. First, a stationary VAR(2) model is used to recover the dynamic responses of the model variables to three structural innovations: the monetary policy, the shock to the growth of neutral productivity, and the shock to the growth rate of investment-specific technology. The sample extends from 1951Q1 to 2008Q4. The set of VAR variables include the growth rate of the relative price of investment, the growth rate of the real GDP-to-hours ratio, inflation, the unemployment rate, capacity utilization, the log of hours, the log of real GDP-to-hours ratio minus the log of real wages, the log of the nominal consumption-to-nominal GDP ratio, the log of the nominal investment-to-nominal GDP ratio, vacancies, the job separation rate, the job finding rate, the log of hours-to-labor force ratio, and the fed funds rate.[3]

Second, the structural parameters in the model are estimated by minimizing the distance between the DSGE model's impulse responses to the structural shocks and the corresponding structural VAR responses. There are three identifying assumptions imposed on the structural VAR model. First, the only variable that the monetary policy shock affects contemporaneously is the federal funds rate. Second, the only shocks that affect labor productivity in the long run are the two technology shocks. Third, the only shock that affects the price of investment relative to consumption in the long run is the innovation to the investment-specific shock. All these identifying assumptions are satisfied in the underlying DSGE model as well. We follow the literature in matching the responses of nine of the fourteen model variables to each of the identified structural shocks.

As is standard in the literature, in estimating θ a subset of the structural parameters is set to values typically imposed in the literature and treated as known: The capital share is $\alpha = 0.25$; the depreciation rate is $\delta = 0.025$; the discount factor is $\beta = 0.999$; the steady-state gross inflation is $\pi = 1.0083$; the government consumption to GDP ratio is 0.2; the

[3]For additional details on the estimation of the structural VAR, the reader may consult Altig et al. (2011) and Christiano et al. (2011). Note that Christiano et al. (2011) treat the federal funds rate as I(0) in their analysis.

relative price of capital in steady state is 1; the wage indexation parameter is $\iota_w = 1$; the wage markup is $\lambda_w = 1.01$; wage stickiness $\xi_w = 0.75$; the gross neutral technology growth is $\mu_z = 1.0041$; and the gross investment technology growth is $\mu_\Psi = 1.0018$.

We consider two alternative estimation methods. The first method corresponds to the classical impulse matching approach as outlined in Rotemberg and Woodford (1997) and Christiano, Eichenbaum, and Evans (2005). The second method corresponds to the Bayesian impulse matching framework recently proposed by Christiano, Trabandt and Walentin (2011). All results shown are based on the diagonal weighting matrix.

4.5.1 Results

Table 3 presents the estimates of the structural parameters based on alternative choices of the maximum horizon of the impulse response functions. We evaluate the structural impulse responses at horizons $0, 1, ..., H$ with $H \in \{15, 19\}$. The order condition $q \leq k$, where q is the number of impulse response parameters to be matched and k is the number of VAR model parameters, is violated for $H = 19$, given that the approximating VAR model in this example includes 14 variables and 2 lags. The order condition is satisfied for $H = 15$. Whether the rank condition holds for $H = 15$ is not readily apparent, but our methods are designed to be asymptotically valid regardless.

The column labelled *Frequentist* reports the results from the impulse response matching estimator based on the diagonal weighting matrix that performed well in the simulation study. The first column corresponds to the point estimate for each structural parameter and the second column shows the standard error. Given the asymptotic normality of this estimator, it makes sense to base inference on these two summary statistics. Alternatively, one could have reported bootstrap percentile intervals. We focus on the standard errors to conserve space.

The next three columns labelled *Bayesian* report results obtained from the quasi-posterior distribution. The first column shows the mode, median and mean of the quasi-

posterior, respectively, as three alternative point estimates of the structural parameters, whereas the second column provides the corresponding standard errors. All Bayesian results are again based on the diagonal weighting matrix and rely on the sandwich formula of Chernozhukov and Hong (2003).

Finally, the column *CTW* reports the mean and standard deviation of the quasi-posterior distributions, computed as in Christiano, Trabandt, and Walentin (2011). As we proved earlier, the variance of the quasi-posterior cannot be used to estimate the asymptotic variance of the structural parameters for $H = 19$, but these standard deviations are reported for comparison.

The qualitative pattern of the results is similar for all choices of H. Table 3 shows that the point estimates are quite robust across alternative methods. This result is not surprising, as all point estimates reported in these tables are consistent. In contrast, there is strong evidence that computing asymptotic standard errors based on the standard deviation of the quasi-posterior tends to understate the uncertainty about the structural parameter estimates. Typically, the standard deviation obtained by the other methods is substantially higher, sometimes by a factor of 3. For example, the standard error of the price markup for $H = 15$ increases from 0.08 to 0.22. Whereas the confidence interval obtained by adding ± 1.96 standard deviations of the quasi-posterior to the point estimate reported in the *CTW* column does reject the null hypothesis that there is no markup (i.e., $\lambda_p = 1$), 95% confidence intervals based on the point estimates and standard errors in the *Frequentist* and *Bayesian* columns do not. Likewise, the standard error of the consumption habit parameter, b, increases threefold compared with the asymptotically invalid estimate reported in the *CTW* column. For $H = 15$, it increases from 0.02 to 0.07. Similarly, the policy reaction function parameters are estimated very imprecisely. For example, ϕ_π is no longer statistically significantly different from zero when conducting inference based on the *Bayesian* intervals, although it does remain significantly different from zero using the *Frequentist* interval. The fact that Bayesian priors affect the results is not surprising, given

the high dimensionality of the model. For $H = 19$, broadly similar results are obtained.

A parameter of particular interest is the price stickiness parameter. The length of the price contracts is defined as $1/(1 - \xi_p)$ quarters, where ξ_p is the probability of not reoptimizing prices today. There is an active literature on measuring the degree of price rigidity at the micro level (see, e.g., Klenow and Kryvtsov (2008), Nakamura and Steinsson (2008)). For example, Klenow and Kryvtsov (2008) provide evidence that price contracts last, on average, about 2.3 quarters. Based on the point estimates for $H = 15$ in Table 3, a researcher would have concluded that the length of a price contract is 2.94, 2.65 or 2.64, respectively. The 95% confidence interval constructed from the information in the CTW column ranges from 2.23 to 3.25 and includes the value of 2.3 from the micro literature. The 95% confidence interval for the length of the price contract implied by the *Frequentist* estimates ranges from 2.12 to 4.79 and includes that value as well, as do the corresponding three 95% confidence intervals implied by the entries in the *Bayesian* columns of Table 3.

This pattern of results changes for $H = 19$. In the latter case, $q > k$, and the conventional asymptotic theory for the CMD estimator breaks down. The results in the CTW column of Table 3 imply a 95% confidence interval with a lower bound of 2.67 that excludes 2.3 for $H = 19$, whereas the three asymptotically valid intervals computed based on the entries in the *Bayesian* columns all include the value of 2.3, suggesting that the DSGE model estimate is consistent with the micro evidence. These estimates are all based on the same prior and hence directly comparable. This example illustrates that the choice of estimation method may affect one's views of whether the macroeconomic evidence is compatible with the length of the price spells found in the micro literature. In contrast, the corresponding 95% confidence interval based on the *Frequentist* method does not include 2.3 for $H = 19$, again illustrating the influence of the prior on the estimates.

The results in Table 3 are based on the conventional premise in empirical work that the structural parameters of interest are strongly identified. The sensitivity of the results to the prior is an indication that the structural parameters may not be strongly identified. Table

4 presents an alternative set of results that allows for the possibility that at least some parameters are only weakly identified. These results are of particular interest for applied work, as none of the currently available methods of inference allows for weak identification of the structural parameters, which is a common problem in the estimation of DSGE models in practice, including the type of DSGE model considered here. We again focus on $H = 16$ and $H = 20$, for expository purposes. Of particular interest is a comparison with the *Frequentist* results in Table 3.

Table 4 shows the lower and upper endpoints of the pointwise 95% interval for each parameter. Allowing for weak identification can affect the substantive conclusions. For example, for both choices of H the null hypothesis that there is no markup (i.e., $\lambda_p = 1$) is rejected. This result differs from the *Frequentist* results in Table 3, which did not allow to reject this null hypothesis for either H. On the other hand, the results for the length of the price contract are qualitatively consistent with *Frequentist* results in Table 3 that were obtained under the premise of strong identification. Whereas the confidence interval for $H = 15$ in Table 4 includes 2.3 quarters, the confidence interval for $H = 19$ in Table 4 does not.

5 Concluding Remarks

One of the leading methods of estimating the structural parameters of DSGE models is the VAR-based impulse response matching estimator. The existing asymptotic theory for this estimator does not cover situations in which the number of impulse response parameters (q) exceeds the number of VAR model parameters (k). Such situations often arise in applied work. We established the consistency of the impulse response matching estimator in this situation, derived its asymptotic distribution and showed how this distribution can be approximated by bootstrap methods. Our results provide formal guidance on how to conduct inference about structural parameters in DSGE models. We also discussed implications of these results for tests of overidentifying restrictions.

Our analysis shed new light on the choice of the weighting matrix. On the one hand, we showed that the impulse response matching estimator based on the optimal weighting matrix, while remaining consistent under our assumptions, does not have an asymptotic normal distribution. Approximating its distribution requires suitable bootstrap methods in practice. On the other hand, we provided a formal justification for the use of bootstrap methods in conducting inference about impulse response matching estimators based on the diagonal weighting matrix. The distribution of the latter estimator was shown to be asymptotically normal, but with a nonstandard asymptotic variance. This result is important because this estimator to date has been used without a formal asymptotic justification having been provided for the practically relevant case of $q > k$. Our analysis also showed that special care is required to ensure that Bayesian methods of inference remain valid from an asymptotic point of view when $q > k$.

We compared the finite-sample accuracy of impulse response matching estimators based on alternative weighting matrices by simulation. A Monte Carlo study based on a small-scale New Keynesian macroeconomic model suggested that the proposed point and interval estimators based on the diagonal weighting matrix are reasonably accurate in finite samples and robust to the choice of the approximating VAR model and the horizon. In contrast, the estimator based on the optimal weighting matrix can be sensitive to the VAR model specification, unless additional finite-sample VAR misspecifications corrections are made. Even in the latter case, its accuracy is slightly lower on average than that of the estimator based on the diagonal weighting matrix.

There is another reason for favoring the impulse response matching estimator based on the diagonal weighting matrix. The condition that $q \leq k$, on which the conventional asymptotic analysis of the impulse response estimator is based, is an order condition, which is only necessary, but not sufficient for the conventional asymptotics to apply. There are cases in which our alternative asymptotic analysis is required even when $q \leq k$ because the rank condition fails. Verifying the rank condition for the use of the delta method is often

difficult in more complicated models. Our analysis showed that verifying this rank condition is not required in practice because, as long as one uses the diagonal weighting matrix, the same bootstrap methods may be used to approximate the asymptotic distribution of the impulse response matching estimator and to compute the asymptotic standard errors for the structural parameters, whether the rank condition holds or not. A similar robustness result is obtained for Bayesian impulse response estimators, provided the sandwich formula of Chernozhukov and Hong (2003) is used.

Finally, we extended the analysis to cover weakly identified DSGE model parameters. Weak identification is a pervasive problem in medium-scale DSGE models. Although several solutions to this problem have been proposed in the recent literature, none apply to impulse response matching estimators. Thus, our analysis greatly extends the range of applications of the impulse response matching estimator. We showed that robustness to weak identification may be achieved by inverting the Wald test statistic of the structural impulse responses to form a joint confidence set and by applying the projection method to recover confidence intervals for individual structural parameters. The proposed method remains asymptotically valid whether $q > k$ or $q \leq k$.

We illustrated the use of the various new methods proposed in this paper in practice. When estimating a prototypical medium-scale New Keynesian DSGE model based on Christiano, Trabandt and Walentin (2011), inference based on the alternative methods proposed in this paper generated substantively different conclusions than methods based on the standard deviation of the quasi-posterior distribution of the structural parameters. For example, whereas the latter method suggested that the macro evidence is inconsistent with micro evidence on the degree of price stickiness at conventional significance levels, given a horizon of 20 quarters, we demonstrated that this result is overturned when using asymptotically valid Bayesian methods of inference. Substantively different results regarding the degree of price stickiness were obtained using frequentist methods, which suggested that the macroeconomic estimates based on a horizon of 20 quarters are inconsistent with the micro

evidence. The latter result was shown to be robust to allowing for weak identification.

Appendix

Proof of Theorem 1.

It follows from the definitions of S_j and Assumption (b) that

$$\Upsilon_T S'(\gamma(\bar{X}_T) - f(\theta)) = \begin{bmatrix} S_0' B_0 Z_T + T^{\frac{1}{2}} S_0'(f(\theta_0) - f(\theta)) \\ S_1' B_1 (Z_T \otimes Z_T) + T S_1'(f(\theta_0) - f(\theta)) \\ \vdots \\ S_r' B_r (Z_T \otimes \cdots \otimes Z_T) + T^{\frac{r+1}{2}} S_r'(f(\theta_0) - f(\theta)) \end{bmatrix} + o_p(1), \quad (45)$$

where $o_p(1)$ is uniform in $\theta \in \Theta$ due to the continuity of $f(\cdot)$ and the compactness of Θ. Because of Assumptions (a) and (b) and because of the continuity of eigenvalues and eigenvectors as a function of matrices, we have

$$\Upsilon_T S \quad (46)$$

$$'(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T)) \xrightarrow{d} \begin{bmatrix} S_0' B_0 Z^* \\ S_1' B_1 (Z^* \otimes Z^*) \\ \vdots \\ S_r' B_r (Z^* \otimes \cdots \otimes Z^*) \end{bmatrix} = \xi^*, \quad (47)$$

where the convergence is with respect to the bootstrap probability measure conditional on the data with probability one. It follows from (46) and Assumption (c) that

$$\Upsilon_T S' \widehat{\Sigma}_T^* S \Upsilon_T = E^*(\xi^* \xi^{*\prime}) + o_p^*(1) = J + o_p^*(1), \quad (48)$$

where $E^*(\cdot)$ is the expectation operator with respect to the bootstrap probability measure conditional on data. It follows from (45), (48) and Assumptions (d) that the objective

function for $\widehat{\theta}_{opt,T}$ is asymptotically proportional to

$$\begin{aligned} & \frac{1}{T^{r+1}}(\widehat{\gamma}_T - f(\theta))'\widehat{\Sigma}_T^{*-1}(\widehat{\gamma}_T - f(\theta)) \\ = \; & \frac{1}{T^{r+1}}(\widehat{\gamma}_T - f(\theta))'S\Upsilon_T \left(\Upsilon_T S'\widehat{\Sigma}_T^* S\Upsilon_T\right)^{-1} \Upsilon_T S'(\widehat{\gamma}_T - f(\theta)) \\ = \; & (f(\theta_0) - f(\theta))'S_r J^{rr-1} S_r'(f(\theta_0) - f(\theta)) + o_p(1) + o_p^*(1), \end{aligned} \tag{49}$$

where J^{rr-1} is the bottom-right $q_r \times q_r$ submatrix of J^{-1} and $o_p(1)$ is uniform in $\theta \in \Theta$. Therefore the consistency of $\widehat{\theta}_{opt,T}$ follows from (49), Assumption (e), and the assumption that $\theta = \theta_0$ is the only solution to $S_r'(f(\theta_0) - f(\theta)) = 0_{q_r \times 1}$. This completes the proof of part (a) of Theorem 1.

To prove part (b), observe that the objective function for $\widehat{\theta}_{diag,T}$ is proportional to

$$(\widehat{\gamma}_T - f(\theta))'W_T(\widehat{\gamma}_T - f(\theta)) \; = \; (f(\theta_0) - f(\theta))'W(f(\theta_0) - f(\theta)) + o_p(1), \tag{50}$$

from which the consistency of $\widehat{\theta}_{diag,T}$ follows.

Proof of Theorem 2. It follows from the first-order conditions and the mean value theorem that

$$\widehat{\theta}_{opt,T} - \theta_0 \; = \; (F(\widehat{\theta}_{opt,T})'\widehat{\Sigma}_T^{*-1}F(\widetilde{\theta}_{opt,T}))^{-1}F(\widehat{\theta}_{opt,T})'\widehat{\Sigma}_T^{*-1}(\widehat{\gamma}_T - f(\theta_0)), \tag{51}$$

$$\widehat{\theta}_{diag,T} - \theta_0 \; = \; (F(\widehat{\theta}_{diag,T})'W_T F(\widetilde{\theta}_{diag,T}))^{-1}F(\widehat{\theta}_T)'W_T(\widehat{\gamma}_T - f(\theta_0)), \tag{52}$$

where $\widetilde{\theta}_{opt,T}$ and $\widetilde{\theta}_{diag,T}$ are points between $\widehat{\theta}_{opt,T}$ and θ_0 and between $\widehat{\theta}_{diag,T}$ and θ_0, respec-

tively, implied by the mean value theorem.

It follows from (45) with $\theta = \theta_0$ and Assumption (a) that

$$\Upsilon_T S'(\gamma(\bar{X}_T) - \gamma(\mu)) \xrightarrow{d} \begin{bmatrix} S_0' B_0 Z \\ S_1' B_1 (Z \otimes Z) \\ \vdots \\ S_r' B_r (Z \otimes \cdots \otimes Z) \end{bmatrix} \equiv \xi. \tag{53}$$

It follows from (48), Theorem 1(a), and Assumptions (d) and (e) that

$$\begin{aligned} & F(\widehat{\theta}_{opt,T})' \widehat{\Sigma}_T^{*-1} F(\widetilde{\theta}_{opt,T}) \\ = \; & F(\widehat{\theta}_{opt,T})' S \Upsilon_T (\Upsilon_T S' \widehat{\Sigma}_T^* S \Upsilon_T)^{-1} \Upsilon_T S' F(\widetilde{\theta}_{opt,T}) \\ = \; & [T^{\frac{1}{2}} F_0' S_0 \; T F_0' S_1 \; \cdots \; T^{\frac{r+1}{2}} F_0' S_r] J^{-1} \begin{bmatrix} T^{\frac{1}{2}} S_0' F_0 \\ T S_1' F_0 \\ \vdots \\ T^{\frac{r+1}{2}} S_r' F_0 \end{bmatrix} + o_p(T^{r+1}) \\ = \; & T^{r+1} F_0' S_r J^{rr-1} S_r' F_0 + o_p(T^{r+1}), \end{aligned} \tag{54}$$

and

$$\begin{aligned} & F(\widehat{\theta}_{opt,T})' \widehat{\Sigma}_T^{*-1} S'^{-1} \Upsilon_T^{-1} \\ = \; & F(\widehat{\theta}_{opt,T})' S \Upsilon_T (\Upsilon_T S' \widehat{\Sigma}_T^* S \Upsilon_T)^{-1} \\ = \; & [T^{\frac{1}{2}} F_0' S_0 \; T F_0' S_1 \; \cdots \; T^{\frac{r+1}{2}} F_0' S_r] J^{-1} + o_p(T^{\frac{r+1}{2}}). \end{aligned} \tag{55}$$

Combining (51), (53), (54) and (55), we obtain

$$T^{\frac{r+1}{2}} (\widehat{\theta}_{opt,T} - \theta_0) \xrightarrow{d} (F_0' S_r J^{rr-1} S_r' F_0)^{-1} F_0' S_r J^{r-1} \xi.$$

Repeating similar arguments we obtain

$$T^{\frac{r+1}{2}}(\widehat{\theta}^*_{opt,T} - \widehat{\theta}_{opt,T}) \xrightarrow{d} (F_0' S_r J^{rr-1} S_r' F_0)^{-1} F_0' S_r J^{r-1} \xi^*,$$

which completes the proof of part (a).

Because

$$T^{\frac{1}{2}}(\widehat{\gamma}_T - f(\theta_0)) = B_0 Z_T + o_p(1), \tag{56}$$

by Assumption (b), it follows from (53) that

$$T^{\frac{1}{2}}(\widehat{\theta}_{diag,T} - \theta_0) \xrightarrow{d} (F_0' W F_0)^{-1} F_0' W B_0 Z, \tag{57}$$

$$T^{\frac{1}{2}}(\widehat{\theta}^*_{diag,T} - \widehat{\theta}_{diag,T}) \xrightarrow{d} (F_0' W F_0)^{-1} F_0' W B_0 Z^*, \tag{58}$$

which completes the proof of part (b).

Proof of Theorem 3.

It follows from the mean value theorem, (46), and Theorem 2(a) that

$$\begin{aligned}
& \widehat{\Sigma}_T^{*-\frac{1}{2}}(\widehat{\gamma}_T - f(\widehat{\theta}_{opt,T})) \\
= \; & (\Upsilon_T S' \widehat{\Sigma}_T^* S \Upsilon_T)^{-\frac{1}{2}} \Upsilon_T S' (\widehat{\gamma}_T - f(\widehat{\theta}_{opt,T})) \\
= \; & (\Upsilon_T S' \widehat{\Sigma}_T^* S \Upsilon_T)^{-\frac{1}{2}} \Upsilon_T S' (\widehat{\gamma}_T - f(\theta_0)) - (\Upsilon_T S' \widehat{\Sigma}_T^* S \Upsilon_T)^{-\frac{1}{2}} \Upsilon_T S' F(\widetilde{\theta}_{opt,T})(\widehat{\theta}_{opt,T} - \theta_0) \\
= \; & J^{-\frac{1}{2}} \xi - J^{-\frac{1}{2}} S_r' F_0 (F_0' S_r J^{rr-1} S_r' F_0)^{-1} F_0' S_r J^{r-1} \xi,
\end{aligned} \tag{59}$$

where $\tilde{\theta}_{opt,T}$ is a point between $\widehat{\theta}_{opt,T}$ and θ_0 and $A^{\frac{1}{2}}$ is the matrix such that $A^{\frac{1}{2}} A^{\frac{1}{2}} = A$. Thus it follows from (59) that

$$J_T \xrightarrow{d} \eta' \eta. \tag{60}$$

The bootstrap version of this result can be derived in a similar way.

Proof of Theorem 4.

We follow the steps in the proof of Theorem 1 in Chernozhukov and Hong (2003) by showing that

$$\int_{H_T} \|h\|^{\alpha} |p_T^*(h) - p_\infty^*(h)| dh \;=\; o_p(1), \tag{61}$$

for every $\alpha \geq 0$, from which Theorem 4 immediately follows. It is convenient to write the localized quasi-posterior as

$$\begin{aligned} p_T^*(h) &= p_T\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) \\ &= \frac{\pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp\left(-q_T(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0))\right)}{\int_{H_T}\pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp\left(-q_T(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0))\right)dh} \\ &= \frac{\pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp(\omega(h))}{\int_{H_T}\pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp(\omega(h))dh}, \end{aligned} \tag{62}$$

where

$$\omega(h) \;=\; -q_T\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) + q_T(\theta_0) + \frac{1}{2}\nabla q_T(\theta_0)'\left(\nabla^2 q_T(\theta_0)\right)^{-1}\nabla q_T(\theta_0).$$

To prove (61), we first show that

$$\int_{H_T}\|h\|^{\alpha}\left|\pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp(\omega(h)) - \pi(\theta_0)\exp\left(-\frac{1}{2}h'F_0'S_r J^{rr-1}S_r'F_0 h\right)\right| dh = o_p(1). \tag{63}$$

Using the second-order Taylor series approximation, $\omega(h)$ can be written as

$$\begin{aligned} \omega(h) &= -\nabla q_T(\theta_0)'\left(\frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) \\ &\quad -\frac{1}{2}\left(\frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)'\nabla^2 q_T(\theta_0)\left(\frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) \\ &\quad -\frac{1}{2}\nabla q_T(\theta_0)'\nabla^2 q_T(\theta_0)]^{-1}\nabla q_T(\theta_0) + R_T\left(\frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) \\ &= -\frac{1}{2T^{r+1}}h'\nabla^2 q_T(\theta_0)h + R_T\left(\frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right), \end{aligned} \tag{64}$$

where $R_T(\cdot)$ is the remainder term. Note that the integral in (63) can be written as the sum of three integrals over (i) $\{h \in H_T : \|h\| \leq M\}$, (ii) $\{h \in H_T : M \leq \|h\| \leq \delta T^{r+1}\}$ and (iii) $\{h \in H_T : \|h\| \geq \delta T^{r+1}\}$. We evaluate each of the three integrals in turn.

Start with the integral over set (i). It follows from the continuity of $\pi(\cdot)$, the smoothness of $f(\cdot)$, (54) and (55) that

$$\sup_{\|h\|\leq M} \left| \pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) - \pi(\theta_0)\right| \xrightarrow{p} 0, \tag{65}$$

$$\sup_{\|h\|\leq M} \left| R_T\left(\frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\right| \xrightarrow{p} 0, \tag{66}$$

from which it follows that

$$\sup_{\|h\|\leq M} \|h\|^\alpha \left| \pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp(\omega(h)) - \pi(\theta_0)\exp\left(-\frac{1}{2}h'\nabla^2 q_T(\theta_0)h\right)\right| = o_p(1). \tag{67}$$

Hence, we have that for every $0 < M < \infty$ and every $\varepsilon > 0$,

$$\begin{aligned} &\liminf_T P_*\left(\int_{h\in H_T:\|h\|\leq M} \|h\|^\alpha \left| \pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp(\omega(h))\right.\right. \\ &\left.\left. -\pi(\theta_0)\exp\left(-\frac{1}{2}h'F_0'S_r J^{rr-1}S_r'F_0 h\right)\right|\right) \\ \geq\ & 1-\varepsilon. \end{aligned} \tag{68}$$

Thus the integral over set (i) is $o_p(1)$.

Next consider the integral over set (ii). Note that

$$\int_{M<\|h\|<\delta T^{\frac{r+1}{2}}} \left|\pi(\theta_0)\exp\left(-\frac{1}{2}h'F_0'S_r J^{rr-1}S_r'F_0 h)\right)\right| dh \tag{69}$$

can be made arbitrarily small by choosing sufficiently large M. Using the quadratic ap-

proximation of $\omega(h)$, (64), we can write

$$\exp(\omega(h)) \;\leq\; \exp\left(-\frac{1}{2}h'\nabla^2 q_T(\theta_0)h + \left|R_T\left(\frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\right|\right) \tag{70}$$

Because f is twice continuously differentiable, Assumption 4(iv)(a) of Chernozhukov and Hong (2003) is satisfied. Thus, for every $\varepsilon > 0$, there are some $\delta > 0$ and $M > 0$ such that

$$\liminf P_*\left(\sup_{M\leq\|h\|\leq\delta T^{\frac{r+1}{2}}} \frac{\left|R_T\left(\frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\right|}{\|h - T^{\frac{r+1}{2}}(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\|^2} \;\leq\; \frac{1}{4}\text{maxeig}(\nabla^2 q_T(\theta_0))\right) \geq\; 1-\varepsilon. \tag{71}$$

Because $(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0) = O_p(1)$, it follows from (70) and (71) that there is C such that

$$\liminf_T P_*\left(\exp(\omega(h)) \leq C\exp\left(-\frac{1}{4}h'\nabla^2 q_T(\theta_0)h\right)\right) \;\geq\; 1-\varepsilon. \tag{72}$$

Combining these results, it follows that

$$\liminf_T P_*\left(\int_{h\in H_T: M<\|h\|<\delta T^{\frac{r+1}{2}}} \|h\|^\alpha \left|\pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp(\omega(h))dh < \varepsilon\right|\right) \;\geq\; 1-\varepsilon \tag{73}$$

from which we obtain

$$\begin{aligned}&\liminf_T P_*\left(\int_{h\in H_T: M<\|h\|<\delta T^{\frac{r+1}{2}}} \|h\|^\alpha \left|\pi(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0))\exp(\omega(h))\right.\right.\\ &\left.\left.-\pi(\theta_0)\exp\left(-\frac{1}{2}h'\nabla^2 q_T(\theta_0)h\right)\right|\right)\\ \geq\;& 1-\varepsilon.\end{aligned} \tag{74}$$

Thus, the integral over set (ii) is also asymptotically negligible.

Third, consider the integral over set (iii). As in the second integral,

$$\int_{\|h\|\geq\delta T^{\frac{r+1}{2}}} \left|\pi(\theta_0)\exp\left(-\frac{1}{2}h'F_0'S_r J^{rr-1}S_r'F_0 h)\right)\right| dh \tag{75}$$

goes to zero on set (iii). Note that

$$\int_{h\in H_T:\|h\|\geq\delta T^{\frac{r+1}{2}}} \|h\|^{\alpha}\exp(\omega(h))\pi\left(\theta_0+\frac{h}{T^{\frac{r+1}{2}}}-(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) \;=\; o_p(1) \qquad (76)$$

is bounded by

$$T^{\frac{(r+1)\alpha}{2}+1}\int_{\|\theta-\theta_0+(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\|\geq\delta}\|\theta-\theta_0+(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\|^{\alpha}\pi(\theta)$$
$$\times\exp\left(q_T(\theta_0)-q_T(\theta)-\frac{1}{2}\nabla q_T(\theta_0)'(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)d\theta. \qquad (77)$$

Since

$$(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0) \;\xrightarrow{p}\; 0, \qquad (78)$$

(77) is in turn bounded by

$$C\exp\left(-\frac{1}{2}\nabla q_T(\theta_0)'(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)T^{1+\frac{\alpha(r+1)}{2}}\int_{\|\theta-\theta_0\|\geq\delta}(1+\|\theta\|^{\alpha})\pi(\theta)\exp(q_T(\theta_0)-q_T(\theta))d\theta. \qquad (79)$$

for some $C>0$. It follows from equation (49) that there is $\varepsilon>0$ such that

$$\liminf_{T\to\infty} P_*\left(\sup_{\|\theta-\theta_0\|\geq\delta/2}\exp(q_T(\theta_0)-q_T(\theta))\leq\exp(-T\epsilon)\right) \;=\; 1. \qquad (80)$$

Hence with probability approaching one (79) is bounded by

$$C\exp\left(-\frac{1}{2}\nabla q_T(\theta_0)'(\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)T^{1+\frac{\alpha(r+1)}{2}}\exp(-T\varepsilon)\int_{\Theta}\|\theta\|^{\alpha}\pi(\theta)d\theta \;=\; o_p(1). \qquad (81)$$

Thus we have

$$\liminf_T P_* \left\{ \int_{h\in H_T:\|h\|\geq \delta T} \|h\|^\alpha \exp(\omega(h))\pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right) \right. \tag{82}$$

$$\left. - \exp\left(-\frac{1}{2}h'F_0'S_rJ^{rr-1}S_r'F_0h)\right)\pi(\theta_0)\right| dh < \varepsilon \Bigg) \;\geq\; 1-\varepsilon \tag{83}$$

In other words, the integral over set (iii) also converges in probability to zero.

Equation (63) follows from (68), (74) and (83). It follows from (63) with $\alpha = 0$ that

$$\begin{aligned}
\int_{H_T} D_T dh &\equiv \pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp(\omega(h))dh \\
&= \int_{H_T} \pi(\theta_0)\exp\left(-\frac{1}{2T^2}h'\nabla^2 q_T(\theta_0)h\right)dh + o_p(1) \\
&= \int_{H_T} \pi(\theta_0)\exp\left(-\frac{1}{2}h'F_0'S_rJ^{rr-1}S_r'F_0h\right)dh + o_p(1) \\
&= \int_{\Re^p} \pi(\theta_0)\exp\left(-\frac{1}{2}h'F_0'S_rJ^{rr-1}S_r'F_0h\right)dh + o_p(1) \\
&= \pi(\theta_0)(2\pi)^{\frac{l}{2}}\left|F_0'S_rJ^{rr-1}S_r'F_0\right|^{-\frac{1}{2}} + o_p(1).
\end{aligned} \tag{84}$$

Thus it follows from (63) and (84) that

$$\begin{aligned}
&\int_{H_T} \|h\|^\alpha |p_T^*(h) - p_\infty^*(h)|dh \\
=\;& \frac{1}{D_T}\int_{H_T}\|h\|^\alpha \left|\pi\left(\theta_0 + \frac{h}{T^{\frac{r+1}{2}}} - (\nabla^2 q_T(\theta_0))^{-1}\nabla q_T(\theta_0)\right)\exp(\omega(h))\right. \\
&\left. - D_T\left(\frac{|F_0'S_rJ^{rr-1}S_r'F_0|}{(2\pi)^l}\right)^{\frac{1}{2}}\exp\left(-\frac{1}{2}h'F_0'S_rJ^{rr-1}S_r'F_0h\right)\right| dh \\
=\;& o_p(1),
\end{aligned} \tag{85}$$

from which (61) follows. This completes the proof of Theorem 4.

Proof of the Proposition. The proof follows from results in Inoue and Kilian (2014),

but is included here for completeness. Recall the definition of S_j and S. It follows from (53) that the limiting covariance matrix of $\Upsilon_T S'(\gamma(\bar{X}_T) - \gamma(\mu))$ is

$$J \;=\; E\left\{ \begin{bmatrix} S_0' B_0 Z \\ S_1' B_1 (Z \otimes Z) \\ \vdots \\ S_r' B_r (Z \otimes \cdots \otimes Z) \end{bmatrix} \begin{bmatrix} S_0' B_0 Z \\ S_1' B_1 (Z \otimes Z) \\ \vdots \\ S_r' B_r (Z \otimes \cdots \otimes Z) \end{bmatrix}' \right\}. \tag{86}$$

Similarly, the limiting covariance matrix of $\Upsilon_T S'(\gamma(\bar{X}_T^*) - \gamma(\bar{X}_T))$ is also (86) conditional on the data with probability one. It follows from Assumption (c) that

$$\Upsilon_T S' \widehat{\Sigma}_T^* S \Upsilon_T \;\xrightarrow{p}\; J. \tag{87}$$

The first part of the Proposition follows from Theorem 1, (53) and (87). The proof of the second part is analogous.

Acknowledgements

We thank Mathias Trabandt for providing access to the data and code used in Christiano et al. (2011). We also have benefitted from helpful discussions with Frank Schorfheide. The views expressed in this paper are those of the authors and do not necessarily reflect those of the Federal Reserve Bank of Philadelphia or of the Federal Reserve System.

References

1. Altig, D., L.J. Christiano, M. Eichenbaum and J. Lindé (2011), "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," *Review of Economic Dynamics* 14, 225-247.

2. Andrews, D.W.K. (1987), "Asymptotic Results for Generalized Wald Tests," *Econometric Theory,* 3, 348-358.

3. Andrews, I. and A. Mikusheva (2015), "Maximum Likelihood Inference in Weakly Identified Models," *Quantitative Economics*, 6, 123-152.

4. Antoine, B., and E. Renault (2012), "Efficient Minimum Distance Estimation with Multiple Rates of Convergence," *Journal of Econometrics,* 1709, 350-367.

5. Bao, Y. (2007), "The Approximate Moments of the Least Squares Estimator for the Stationary Autoregressive Model Under a General Error Distribution," *Econometric Theory*, 23, 1013-1021.

6. Bao, Y., and A. Ullah (2007), "The Second-Order Bias and Mean Squared Error of Estimators in Time Series Models," *Journal of Econometrics*, 140, 650-669.

7. Benkwitz, A., M.H. Neumann, and H. Lütekpohl (2000), "Problems Related to Confidence Intervals for Impulse Responses of Autoregressive Processes," *Econometric Reviews*, 19, 69-103.

8. Boivin, J., and M. Giannoni (2006), "Has Monetary Policy Become More Effective?", *Review of Economics and Statistics*, 88, 445-462.

9. Canova, F., and L. Sala (2009), "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics,* 56, 431-449.

10. Chauduri, S., and E. Zivot (2011), "A New Method of Projection-Based Inference in GMM with Weakly Identified Nuisance Parameters," *Journal of Econometrics,* 164, 239-251.

11. Chernozhukov, V., and H. Hong (2003), "An MCMC Approach to Classical Estimation," *Journal of Econometrics*, 196, 293-346.

12. Christiano, L.J., M. Eichenbaum and C. Evans (2005), "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113, 1-45.

13. Christiano, L.J., M. Eichenbaum and M. Trabandt (2013), "Unemployment and Business Cycles," unpublished manuscript, Northwestern University and Board of Governors of the Federal Reserve System.

14. Christiano, L.J., M. Eichenbaum and R.J. Vigfusson (2006), "Assessing Structural VARs," *Macroeconomics Annual 2006*, Acemoglu, D., K. Rogoff and M. Woodford (eds.), MIT Press, Boston, MA.

15. Christiano, L.J., M. Trabandt and K. Walentin (2011), "DSGE Models for Monetary Policy Analysis," in: Friedman, B.M. and M. Woodford (eds.): *Handbook of Monetary Economics,* Volume 3A, North-Holland, Amsterdam.

16. DiCecio, R. (2005), "Comovement: It's Not a Puzzle", Working Paper 2005-035, *Federal Reserve Bank of St. Louis*.

17. DiCecio, R., and E. Nelson (2007), "An Estimated DSGE Model for the United Kingdom", *Review*, Federal Reserve Bank of St. Louis, 89, 215-231.

18. Dridi, R., A. Guay, and E. Renault (2007), "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models", *Journal of Econometrics*, 136, 397-430.

19. Dufour, J.M., L. Khalaf, and M. Kichian (2013), "Identification-Robust Analysis of DSGE and Structural Macroeconomic Models," *Journal of Monetary Economics*, 60, 340-350.

20. Dufour, J.M., and M. Taamouti (2005), "Projection-Based Statistical Inference in Linear Structural Models with Possibly Weak Instruments," *Econometrica,* 73, 1351-1365.

21. Dupor, B., J. Han, and Y.C. Tsai (2007), "What Do Technology Shocks Tell Us about the New Keynesian Paradigm?", *Journal of Monetary Economics* 56, 560-569.

22. Fernandez-Villaverde, J., J.F. Rubio-Ramirez, T.J. Sargent and M.W. Watson (2007), "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97, 1021-1026.

23. Guerron-Quintana, P., A. Inoue and L. Kilian (2013), "Frequentist Inference in Weakly Identified DSGE Models," *Quantitative Economics,* 4, 197-227.

24. Hall, P. (1992), *The Bootstrap and Edgeworth Expansion,* Springer: New York.

25. Hall, P., and J.L. Horowitz (1996), "Bootstrap Critical Values for Tests Based on Generalized-Method-of-Moments Estimators," *Econometrica,* 64, 891-916.

26. Hall, A.R., A. Inoue, J.M. Nason, and B. Rossi (2012), "Information Criteria for Impulse Response Function Matching Estimation of DSGE Models," *Journal of Econometrics*, 170, 499-518.

27. Iacoviello, M. (2005), "House Prices, Borrowing Constraints and Monetary Policy in the Business Cycle," *American Economic Review*, 95, 739-764.

28. Inoue, A. and L. Kilian (2002), "Bootstrapping Smooth Functions of Slope Parameters and Innovation Variances in VAR(∞) Models," *International Economic Review*, 43, 309-332.

29. Inoue, A., and L. Kilian (2014), "Joint Confidence Sets for Structural Impulse Responses," manuscript, Vanderbilt University and University of Michigan.

30. Jordà, Ò., and S. Kozicki (2011), "Estimation and Inference by the Method of Projection Minimum Distance," *International Economic Review,* 52, 461-487.

31. Kim, J.Y. (2002), "Limited Information Likelihood and Bayesian Analysis," *Journal of Econometrics*, 107, 175-193.

32. Klenow, P. J. and O. Kryvtsov (2008), "State-Dependent or Time-Dependent Pricing: Does It Matter for Recent U.S. Inflation?," *Quarterly Journal of Economics*, 123, 863-904.

33. Kormilitsina, A., and D. Nekipelov (2013), "Consistent Variance of the Laplace Type Estimators: Application to DSGE Models," forthcoming: *International Economic Review*.

34. Lütkepohl, H. (1990), "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *Review of Economics and Statistics,* 72, 116-25.

35. Magnus, J.R., and H. Neudecker (1999), *Matrix Differential Calculus with Applications in Statistics and Econometrics,* Revised Edition, Wiley, Chichester.

36. Nakamura, E. and J. Steinsson (2008), "Five Facts about Prices: A Reevaluation of Menu Cost Models," *Quarterly Journal of Economics*, 123, 1415-1464.

37. Newey, W.K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing", in: R.F. Engle and D.L. McFadden, *Handbook of Econometrics*, Vol. 4, Elsevier.

38. Newey, W.K., and R.J. Smith (2004), "Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators," *Econometrica,* 72, 219-255.

39. Phillips, P.C.B. (1989), "Partially Identified Econometric Models," *Econometric Theory,* 5, 181-240.

40. Qu, Z. (2014), "Inference in Dynamic Stochastic General Equilibrium Models with Possible Weak Identification," *Quantitative Economics*, 5, 457-494.

41. Rotemberg, J. and M.W. Woodford (1997), "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy", in: Bernanke, B., and J. Rotemberg (eds.), *NBER Macroeconomics Annual*, MIT Press, Cambridge, MA.

42. Sims, C.A., J.H. Stock and M.W. Watson (1990), "Inference in Linear Time Series Models with Some Unit Roots," *Econometrica,* 58, 113-144.

43. Smets, F. and R. Wouters (2007), "Shocks and Frictions in U.S. Business Cycles: A Bayesian Approach," *American Economic Review* 97, 586-606.

44. Stock, J.H., and J.H. Wright (2000), "GMM with Weak Identification," *Econometrica,* 68, 1055-1096.

45. Uribe, M., and V.Z. Yue (2006), "Country Spreads and Emerging Markets: Who Drives Whom?", *Journal of International Economics* 69, 6-36.

46. Woodford, M. (2003), *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press, Princeton, NJ.

Table 1: Bias and MSE of the Parameter a in the Small-Scale New Keynesian Model

Baseline Specification									
			Diagonal Weighting Matrix				Optimal Weighting Matrix			
T	p	H	Mean Bias	Median Bias	Variance	MSE	Mean Bias	Median Bias	Variance	MSE
100	6	12	-0.0076	-0.0093	0.0009	0.0009	0.1002	0.0976	0.0020	0.0120
100	6	16	-0.0088	-0.0116	0.0009	0.0010	0.1042	0.1050	0.0020	0.0129
100	8	12	-0.0085	-0.0126	0.0009	0.0010	0.0777	0.0766	0.0015	0.0076
100	8	16	-0.0067	-0.0098	0.0010	0.0010	0.0840	0.0809	0.0017	0.0088
100	10	12	-0.0033	-0.0057	0.0010	0.0010	0.0722	0.0699	0.0011	0.0063
100	10	16	-0.0050	-0.0066	0.0010	0.0010	0.0779	0.0764	0.0014	0.0074
100	12	12	0.0022	-0.0011	0.0012	0.0012	0.0644	0.0639	0.0008	0.0050
100	12	16	0.0023	-0.0004	0.0011	0.0011	0.0792	0.0778	0.0012	0.0075
232	6	12	-0.0224	-0.0236	0.0003	0.0008	0.1201	0.1209	0.0021	0.0008
232	6	16	-0.0208	-0.0223	0.0004	0.0008	0.1254	0.1252	0.0021	0.0178
232	8	12	-0.0211	-0.0226	0.0003	0.0007	0.0687	0.0644	0.0017	0.0064
232	8	16	-0.0199	-0.0212	0.0003	0.0007	0.0760	0.0709	0.0017	0.0075
232	10	12	-0.0193	-0.0196	0.0003	0.0007	0.0510	0.0475	0.0010	0.0036
232	10	16	-0.0188	-0.0198	0.0003	0.0007	0.0536	0.0481	0.0011	0.0040
232	12	12	-0.017	-0.0168	0.0003	0.0006	0.0295	0.0293	0.0004	0.0013
232	12	16	-0.0188	-0.0201	0.0003	0.0007	0.0467	0.0457	0.0007	0.0029

With Finite-Order VAR Misspecification Bias Correction									
			Diagonal Weighting Matrix				Optimal Weighting Matrix			
T	p	H	Mean Bias	Median Bias	Variance	MSE	Mean Bias	Median Bias	Variance	MSE
100	6	12	-0.0373	-0.0415	0.0087	0.0101	-0.0306	-0.0324	0.0072	0.0081
100	6	16	-0.0360	-0.0403	0.0087	0.0099	-0.0261	-0.0310	0.0074	0.0081
100	8	12	-0.0305	-0.0329	0.0084	0.0093	-0.0271	-0.0301	0.0082	0.0089
100	8	16	-0.0292	-0.0323	0.0084	0.0093	-0.0198	-0.0258	0.0086	0.0090
100	10	12	-0.0226	-0.0268	0.0084	0.0089	-0.0251	-0.0293	0.0081	0.0087
100	10	16	-0.0226	-0.0268	0.0084	0.0089	-0.0227	-0.0286	0.0083	0.0088
100	12	12	-0.0252	-0.0284	0.0082	0.0089	-0.0180	-0.0249	0.0078	0.0081
100	12	16	-0.0253	-0.0286	0.0083	0.0090	-0.0090	-0.0203	0.0080	0.0081
232	6	12	-0.0286	-0.0362	0.0086	0.0094	-0.0415	-0.0352	0.0063	0.0080
232	6	16	-0.0287	-0.0366	0.0086	0.0094	-0.0410	-0.0351	0.0063	0.0079
232	8	12	-0.0377	-0.0326	0.0067	0.0081	-0.0373	-0.0315	0.0065	0.0079
232	8	16	-0.0373	-0.0344	0.0070	0.0084	-0.0347	-0.0316	0.0072	0.0084
232	10	12	-0.0368	-0.0350	0.0087	0.0100	-0.0328	-0.0326	0.0069	0.0080
232	10	16	-0.0365	-0.0347	0.0087	0.0100	-0.0315	-0.0320	0.0070	0.0080
232	12	12	-0.0379	-0.0361	0.0081	0.0096	-0.0328	-0.0310	0.0068	0.0079
232	12	16	-0.0397	-0.0363	0.0082	0.0098	-0.0350	-0.0315	0.0076	0.0088

Notes: T denotes the sample size, p the VAR lag order, and H the maximum horizon of the impulse response functions. a is the probability of a firm not adjusting its price.

Table 2: Effective Coverage Probabilities of Confidence Intervals for the Parameter a in the Small-Scale New Keynesian Model

Baseline Specification						
			Diagonal Weighting Matrix		Optimal Weighting Matrix	
T	p	H	68%	90%	68%	90%
100	6	12	67.82	87.80	57.90	68.28
100	6	16	68.00	88.00	55.00	66.60
100	8	12	66.60	86.80	71.00	84.80
100	8	16	66.20	86.60	70.00	83.20
100	10	12	66.00	85.80	66.80	89.20
100	10	16	66.40	85.80	67.40	84.00
100	12	12	64.00	84.40	64.20	87.40
100	12	16	63.80	86.20	67.60	87.80
232	6	12	68.40	91.20	19.00	42.60
232	6	16	65.20	90.80	20.00	39.80
232	8	12	72.60	93.20	74.20	87.00
232	8	16	69.80	73.40	89.00	85.40
232	10	12	66.40	90.20	74.20	91.00
232	10	16	72.60	90.60	80.60	91.40
232	12	12	70.00	90.80	68.60	70.00
232	12	16	70.80	89.60	71.20	93.00

With Finite-Order VAR Misspecification Bias Correction						
			Diagonal Weighting Matrix		Optimal Weighting Matrix	
T	p	H	68%	90%	68%	90%
100	6	12	69.00	87.40	67.60	85.60
100	6	16	68.20	87.40	67.60	84.20
100	8	12	69.00	86.60	64.60	83.40
100	8	16	68.00	86.20	65.00	80.60
100	10	12	68.20	85.60	63.20	82.40
100	10	16	68.20	85.60	63.00	81.60
100	12	12	69.60	87.40	63.80	83.00
100	12	16	69.80	87.00	64.60	81.00
232	6	12	72.20	87.00	71.00	90.20
232	6	16	72.20	87.00	69.80	90.20
232	8	12	74.60	89.40	67.60	90.60
232	8	16	75.00	88.40	67.60	88.80
232	10	12	67.80	87.20	68.80	87.8
232	10	16	68.40	87.20	68.00	86.80
232	12	12	69.80	88.40	68.60	88.00
232	12	16	69.00	88.60	64.60	86.20

Notes: T denotes the sample size, p the VAR lag order, and H the maximum horizon of the impulse response functions. a is the probability of a firm not adjusting its price.

Table 3: Estimates for Medium-Scale DSGE Model Assuming Strong Identification

H = 15		Frequentist		Bayesian						CTW	
		Estimate	S.E.	Mode	S.E.	Median	S.E.	Mean	S.E.	Mean	S.E.
Price Stickiness	ξ_p	0.660	0.068	0.623	0.078	0.623	0.077	0.622	0.077	0.622	0.036
Std. Monetary Policy	$\sigma_{\epsilon,r}$	0.592	0.077	0.504	0.077	0.509	0.077	0.509	0.077	0.509	0.044
Std. Neutral Tech. Shock	$\sigma_{\epsilon,\mu}$	0.228	0.059	0.224	0.056	0.225	0.056	0.226	0.055	0.226	0.016
Autocorr. Invest. Tech. Specific	$\rho_{\mu,\psi}$	0.617	0.105	0.620	0.096	0.607	0.102	0.600	0.104	0.600	0.071
Std. Invest. Tech. Specific	$\sigma_{\epsilon,\psi}$	0.152	0.038	0.158	0.040	0.159	0.040	0.160	0.040	0.160	0.022
Taylor Rule: Interest Smoothing	ρ_r	0.820	0.054	0.880	0.068	0.879	0.068	0.879	0.069	0.879	0.015
Taylor Rule: Inflation	ϕ_π	1.030	0.365	1.449	0.862	1.462	0.870	1.472	0.874	1.472	0.113
Taylor Rule: Output Gap	ϕ_y	0.0005	0.041	0.087	0.187	0.091	0.196	0.095	0.202	0.095	0.033
Investment Adjustment Costs	S"	16.100	8.090	12.435	6.207	13.028	6.563	13.320	6.744	13.320	2.697
Consumption Habit	b	0.757	0.071	0.763	0.064	0.762	0.065	0.761	0.065	0.761	0.018
Capacity Adjustment Costs	σ_a	0.156	0.189	0.296	0.269	0.312	0.286	0.318	0.290	0.318	0.081
Price Markup	λ_p	1.250	0.215	1.178	0.210	1.189	0.217	1.194	0.221	1.194	0.078
Inverse Labor Supply Elasticity	υ	0.146	0.135	0.106	0.092	0.108	0.095	0.111	0.098	0.111	0.023

H=19		Frequentist		Bayesian						CTW	
		Estimate	S.E.	Mode	S.E.	Median	S.E.	Mean	S.E.	Mean	S.E.
Price Stickiness	ξ_p	0.744	0.063	0.711	0.074	0.707	0.076	0.704	0.077	0.704	0.040
Std. Monetary Policy Shock	$\sigma_{\epsilon,r}$	0.596	0.074	0.552	0.078	0.551	0.078	0.551	0.077	0.551	0.043
Std. Neutral Tech. Shock	$\sigma_{\epsilon,\mu}$	0.331	0.090	0.293	0.084	0.293	0.084	0.293	0.084	0.293	0.028
Autocorr. Invest. Tech. Shock	$\rho_{\mu,\psi}$	0.678	0.075	0.663	0.098	0.649	0.103	0.643	0.105	0.643	0.060
Std. Invest. Tech. Shock	$\sigma_{\epsilon,\psi}$	0.143	0.044	0.148	0.045	0.152	0.045	0.154	0.044	0.154	0.021
Taylor Rule: Interest Smoothing	ρ_r	0.868	0.056	0.887	0.075	0.885	0.077	0.885	0.078	0.885	0.014
Taylor Rule: Inflation	ϕ_π	1.260	0.464	1.464	0.890	1.479	0.907	1.488	0.914	1.488	0.110
Taylor Rule: Output Gap	ϕ_y	0.175	0.166	0.188	0.298	0.200	0.319	0.206	0.329	0.206	0.060
Investment Adjustment Costs	S"	6.570	3.114	7.750	4.222	8.069	4.472	8.298	4.639	8.298	1.888
Consumption Habit	b	0.739	0.071	0.748	0.067	0.746	0.069	0.746	0.069	0.746	0.020
Capacity Adjustment Costs	σ_a	0.370	0.270	0.340	0.246	0.355	0.267	0.362	0.276	0.362	0.080
Price Markup	λ_p	1.250	0.218	1.166	0.195	1.174	0.203	1.180	0.207	1.180	0.074
Inverse Labor Supply Elasticity	υ	0.0548	0.051	0.062	0.060	0.066	0.064	0.069	0.066	0.069	0.017

Notes: H denotes the maximum horizon of the impulse response functions.

Table 4: 95% Confidence Intervals for Medium-Scale DSGE Model Allowing for Weak Identification

		$H = 15$		$H = 19$	
Price Stickiness	ξ_p	0.460	0.726	0.579	0.796
Std. Monetary Policy Shock	$\sigma_{\epsilon,r}$	0.380	0.664	0.385	0.690
Std. Neutral Tech. Shock	$\sigma_{\epsilon,\mu}$	0.182	0.299	0.211	0.374
Autocorr. Invest. Tech. Shock	$\rho_{\mu,\psi}$	0.330	0.764	0.437	0.792
Std. Invest. Tech. Shock	$\sigma_{\epsilon,\psi}$	0.108	0.231	0.101	0.239
Taylor Rule: Interest Smoothing	ρ_r	0.840	0.922	0.841	0.921
Taylor Rule: Inflation	ϕ_π	1.210	1.912	1.241	1.915
Taylor Rule: Output Gap	ϕ_y	0.018	0.262	0.063	0.425
Investment Adjustment Costs	S"	6.902	26.245	4.327	18.819
Consumption Habit	b	0.702	0.815	0.681	0.811
Capacity Adjustment Costs	σ_a	0.078	0.683	0.155	0.632
Price Markup	λ_p	1.017	1.502	1.015	1.464
Inverse Labor Supply Elasticity	υ	0.056	0.227	0.027	0.162

Notes: H denotes the maximum horizon of the impulse response functions.