# Inflation dynamics and the New Keynesian Phillips Curve: an identification robust econometric analysis*

Jean-Marie Dufour † Université de Montréal

Lynda Khalaf ‡ Université Laval

Maral Kichian § Bank of Canada

First version: April 2004
Revised: September 2004, April 2005, August 2005
This version: August 2005
Compiled: May 18, 2006, 9:53am

This paper is forthcoming in the *Journal of Economic Dynamics and Control.*

* The views in this paper are our own and do not represent those of the Bank of Canada or its staff. We would like to thank Florian Pelegrin, two anonymous referees and the Editors Jim Bullard, Cees Diks and Florain Wegener for several useful comments. Wendy Chan and Amanda Armstrong also supplied excellent research assistance. This work was supported by the Bank of Canada, the Canada Research Chair Program [Chair in Econometrics, Université de Montréal, and Chair in Environmental and Financial Econometric Analysis, Université Laval], the Institut de Finance mathématique de Montréal (IFM2), the Canadian Network of Centres of Excellence [program on *Mathematics of Information Technology and Complex Systems* (MITACS)], the Alexander-von-Humboldt Foundation (Germany, Konrad-Adenauer Research Award), the Natural Sciences and Engineering Research Council of Canada, the Social Sciences and Humanities Research Council of Canada, the Fonds de recherche sur la société et la culture (Québec), and the Fonds de recherche sur la nature et les technologies (Québec).

† Canada Research Chair Holder (Econometrics). Centre interuniversitaire de recherche en analyse des organisations (CIRANO), Centre interuniversitaire de recherche en économie quantitative (CIREQ), and Département de sciences économiques, Université de Montréal. Mailing address: Département de sciences économiques, Université de Montréal, C.P. 6128 succursale Centre-ville, Montréal, Québec, Canada H3C 3J7. TEL: 1 (514) 343 2400; FAX: 1 (514) 343 5831; e-mail: jean.marie.dufour@umontreal.ca . Web page: http://www.fas.umontreal.ca/SCECO/Dufour

‡ Département d'économique and Groupe de Recherche en économie de l'énergie de l'environement et des ressources naturelles (GREEN), Université Laval, and Centre Interuniversitaire de Recherche en économie quantitative (CIREQ), Université de Montréal. Mailing address: GREEN, Université Laval, Pavillon J.-A.-De Sève, St. Foy, Québec, Canada, G1K 7P4. TEL: (418) 656 2131-2409; FAX: (418) 656 7412; email: lynda.khalaf@ecn.ulaval.ca

§ Research Department, Bank of Canada, Ottawa, Canada K1A 0G9. TEL: (613) 782 7460; FAX: (613) 782 7163; email: mkichian@bankofcanada.ca

**ABSTRACT**

In this paper, we use identification-robust methods to assess the empirical adequacy of a New Keynesian Phillips Curve (NKPC) equation. We focus on the Gali and Gertler's (1999) specification, on both U.S. and Canadian data. Two variants of the model are studied: one based on a rational-expectations assumption, and a modification to the latter which consists in using survey data on inflation expectations. The results based on these two specifications exhibit sharp differences concerning: (i) identification difficulties, (ii) backward-looking behavior, and (iii) the frequency of price adjustments. Overall, we find that there is some support for the hybrid NKPC for the U.S., whereas the model is not suited to Canada. Our findings underscore the need for employing identification-robust inference methods in the estimation of expectations-based dynamic macroeconomic relations.

**Key words**: macroeconomics; inflation dynamics; New Keynesian Phillips Curve; identification robust inference; weak instruments; optimal instruments.

**JEL classification**: C12, C13, C3, C52, E3, E31, E5.

**RÉSUMÉ**

Dans cet article, nous employons des méthodes robustes aux problèmes d'identification afin d'évaluer la valeur empirique d'une nouvelle équation de courbe de Phillips keynésienne (NKPC). Nous concentrons notre analyse sur la spécification de Gali et Gertler (1999), en considérant des données américaines et canadiennes. Nous étudions deux variantes du modèle: une première fondée sur une hypothèse d'attentes rationnelles et une seconde où les attentes sont mesurées à partir de données d'enquête. Les résultats basés sur ces deux spécifications diffèrent de manière notable sur plusieurs points: (i) les problèmes liés à l'identification, (ii) les comportements rétrospectifs, (iii) la fréquence des ajustements. En fin de compte, nos résultats sont compatibles dans une faible mesure avec un modèle NKPC hybride, tandis que le modèle ne semble pas compatible avec les données canadiennes. Nos résultats soulignent l'importance d'utiliser des méthodes robustes à l'identification dans l'analyse empirique de relations macroéconomiques où interviennent des attentes.



**Mots-clefs:** macroéconomie; dynamique de l'inflation; nouvelle courbe de Phillips keynésienne; inférence robuste à l'identification; instruments faibles; instruments optimaux.

**Classification du Journal of Economic Literature:** C12, C13, C3, C52, E3, E31, E5.

## Contents



## List of Tables



## List of Figures

# 1. Introduction

A standard feature of macroeconomic policy models is an equation describing the evolution of inflation. Nowadays, this process is typically modelled as a hybrid New Keynesian Phillips curve (NKPC). This specification results from recent efforts to model the short-run dynamics of inflation starting from optimization principles; see, for example, Woodford (2003) and the references therein. In its basic form, the NKPC stipulates that inflation at time $t$ is a function of expected future inflation and the current output gap. With its clearly-elucidated theoretical foundations, the NKPC possesses a straightforward structural interpretation and therefore presents, in principle, a strong theoretical advantage over traditional reduced-from Phillips curves (which are only statistically justified).

However, given the statistical failure of the basic NKPC formulation when confronted with data, the curve has since evolved into its more empirically-viable hybrid form. In particular, it was noted that: (i) adding lagged inflation to the model (*hybrid NKPC*) corrects the signs of estimated coefficients [see Fuhrer and Moore (1995), Fuhrer (1997) and Roberts (1997)], and (ii) using a measure of real marginal cost derived from a given production function instead of the output gap yields a better statistical fit according to GMM-based estimates and tests [see, for example, Gali and Gertler (1999) and Gali, Gertler and Lopez-Salido (2001)]. Yet the question of which production function (*i.e.*, which marginal cost measure) is empirically preferable is not yet resolved, as the choice for the marginal cost proxy seems to affect evidence on the weight of the backward-looking term; see Gagnon and Khan (2005). In addition, there are different theoretical ways of incorporating backward-looking behavior in the curve, and they yield different outcomes; see Fuhrer and Moore (1995), Gali and Gertler (1999) and Eichenbaum and Fisher (2004).[1]

Discriminating between competing alternatives calls for reliable econometric methods. Full-information models are typically nonlinear and heavily parametrized.[2] So, in practice, these models are often estimated by applying standard limited-information (LI) instrumental-variable (IV) methods to first-order conditions of interest. Indeed, the popularity of NKPC models stems in large part from studies such as Gali and Gertler (1999) and Gali et al. (2001) who found empirical support for their version of the curve using the generalized method of moments (GMM), and the fact that the model is not rejected by Hansen's $J$ test.

But even as the popularity and usage of the curve has grown, criticisms have been raised with respect to its empirical identifiability. The main issue is that IV methods such as GMM are not immune to the presence of weak instruments; see, for example, Dufour (1997, 2003), Staiger and Stock (1997), Wang and Zivot (1998), Zivot, Startz and Nelson (1998), Stock and Wright (2000), Dufour and Jasiak (2001), Stock, Wright and Yogo (2002), Kleibergen (2002), Khalaf and Kichian (2002, 2005), Dufour and Khalaf (2003), and Dufour and Taamouti (2005, 2006, 2003*b*, 2003*a*). These studies demonstrate that standard asymptotic procedures (which *impose identification away* without correcting for local almost-nonidentification) are fundamentally flawed and lead to spurious overrejections, even with fairly large samples. In particular, the following fundamental problems do occur: in models which may not be identified over all the parameter space, (i) usual $t$-type tests

[1] For example, Gali and Gertler (1999) appeal to the assumption that a proportion of firms never re-optimize, but that they set their prices using a rule-of-thumb method; Eichenbaum and Fisher (2004) use dynamic indexing instead.

[2] In this literature, some of the parameters are typically calibrated while others are estimated.

have significance levels that may deviate arbitrarily from their nominal levels since it is not possible to bound the null distributions of the test statistic, and (ii) Wald-type confidence intervals [of the form: estimate $\pm$ (asymptotic standard error) $\times$ (asymptotic critical point)] have dramatically poor coverage irrespective of their nominal level because they are bounded by construction; see Dufour (1997).[3]

To circumvent the difficulties associated with weak instruments, the above cited recent work on IV-based inference has focused on two main directions [see the surveys of Dufour (2003) and Stock et al. (2002)]: (i) refinements in asymptotic analysis which hold whether instruments are weak or not [*e.g.*, Staiger and Stock (1997), Wang and Zivot (1998)), Stock and Wright (2000), Kleibergen (2002), Moreira (2003*b*)], and (ii) finite-sample procedures based on proper pivots, i.e. statistics whose null distributions do not depend on nuisance parameters or can be bounded by nuisance-parameter-free distributions (*boundedly pivotal functions*) [Dufour (1997), Dufour and Jasiak (2001), Dufour and Khalaf (2002), and Dufour and Taamouti (2005, 2006, 2003*b*, 2003*a*)]. The latter include methods based on Anderson and Rubin's (1949, AR) *pivotal $F$*-statistic which allow *unbounded* confidence sets.

Identification difficulties have led to re-examinations of NKPC models, and in particular of the Gali and Gertler NKPC specification, by several authors. Especially relevant contributions on this issue include Linde (2001), Ma (2002), Nason and Smith (2003) and Fuhrer and Olivei (2004). Linde (2001) performs a small-scale simulation study based on a Gali-Gertler-type model and documents the superiority of full-information maximum likelihood (FIML) over GMM. In particular, GMM estimates appear sensitive to parameter calibrations. Ma (2002) applies the asymptotic methods proposed by Stock and Wright (2000) to the Gali and Gertler's NKPC in view of getting confidence sets that account for the presence of weak instruments. These sets are much too large to be informative, suggesting that the parameters of the curve are indeed not well-identified. Nason and Smith (2003) study the identification issue of the NKPC in limited-information contexts analytically, solving the Phillips curve difference equation. They show that typical GMM estimations of such curves have parameters that are not identifiable (or nearly so), and full-information methods (FIML) can make identification easier. Applications to U.S. data yield GMM estimates that are comparable to the values obtained by Gali and Gertler (1999). In contrast, their FIML estimates (which the authors feel are more reliable) point to a greater role for backward-looking behavior. For Canada, the authors report that the NKPC is poorly identified, whether GMM or FIML estimation is used. Finally, Fuhrer and Olivei (2004) consider improved GMM estimation, where the instrumentation stage takes the constraints implied by the structure formally into consideration. They demonstrate the superiority of their approach through a Monte Carlo simulation. In addition, they estimate an inflation equation using U.S. data, and obtain a large forward looking component with conventional GMM, but a much lower value for this parameter with "optimal" GMM and maximum likelihood.

In this paper, we reconsider the problem of estimating inflation dynamics, in view of recent

[3] Poor coverage (which implies that the data is uninformative about the parameter in question) is not really due to large estimated standard errors, or even to poorly approximated cut-off points. The problems stem from the method of building the confidence set as an interval which is automatically "bounded". Any valid method for the construction of confidence sets should allow for possibly unbounded outcomes, when the admissible set of parameter values is unbounded (as occurs when parameters are not identifiable on a subset of the parameter space). In this case, a bounded confidence set would inevitably "rule out" plausible parameter sets, with obvious implications on coverage.

econometric findings. Our aim is to produce more reliable inference based on identification-robust tests and confidence sets. A characteristic feature of identification-robust procedures is they should lead to uninformative (e.g., unbounded) confidence sets when the parameters considered are not identified [see Dufour (1997)]. We focus on two types of procedures: the AR procedure and a method proposed by Kleibergen (2002). The AR procedure is particularly appropriate from the viewpoint of validating a structural model, because it is robust not only to weak instruments, but also to missing instruments and more generally to the formulation of a model for endogenous explanatory variables [see Dufour (2003) and Dufour and Taamouti (2005, 2006)]. A drawback, however, of the AR procedure comes from the fact that it leads to the inclusion of a potentially large number of additional regressors (identifying instruments), hence a reduction in degrees of freedom which can affect test power in finite samples. To assess sensitivity to this type of effect, we also apply a method proposed by Kleibergen (2002), which may yield power gains by reducing the number of "effective" regressors (although at the expense of some robustness).[4]

Our applications study U.S. and Canadian data using: (i) the benchmark hybrid NKPC of Gali and Gertler, which uses a rational expectations assumption, and (ii) a modification to the latter which consists in using survey-based measures of expected inflation. Our analysis allows one to compare and contrast both variants of the model; this is relevant because available studies imply that the specification of the expectation variable matters empirically. For instance, Gali and Gertler (1999) suggest that, when the model is conditional on labour costs, under rational expectations, additional lags of inflation are no longer needed. In contrast, Roberts (2001) argues that those results are sensitive to the specification of labour costs, and that the need to include additional lags could reflect the fact that expectations are not rational; see also Roberts (1997, 1997, 1998). Our results reveal sharp differences between the two specifications for U.S. and Canada.

In section 2, we review the Gali and Gertler's (1999) NKPC hybrid specification. In section 3, we describe the specific models and the methodology used in this paper. Section 4 discusses our empirical results, and section 5 concludes. Details on the data and a formal treatment of the statistical procedures we apply are presented in Appendices A and B.

## 2. Gali and Gertler's hybrid NKPC model

In Gali and Gertler's hybrid specification, firms evolve in a monopolistically competitive environment and cannot adjust their prices at all times. A Calvo-type assumption is used to represent the fact that a proportion $\theta$ of the firms do not adjust their prices in period $t$. In addition, it is assumed that some firms do not optimize but use a rule of thumb when setting their prices. The proportion of such firms (referred to as the backward-looking price-setters) is given by $\omega$. In such an environment, profit-maximization and rational expectations lead to the following hybrid NKPC equation for inflation ($\pi_t$):

$$\pi_t = \lambda s_t + \gamma_f E_t \pi_{t+1} + \gamma_b \pi_{t-1}, \tag{2.1}$$

$$\pi_{t+1} = E_t \pi_{t+1} + \upsilon_{t+1} \tag{2.2}$$

[4] For further discussion of this issue, see Dufour and Taamouti (2003*b*, 2003*a*).

where

$$\lambda = \frac{(1-\omega)(1-\theta)(1-\beta\theta)}{\theta+\omega-\omega\theta+\omega\beta\theta}\,, \tag{2.3}$$

$$\gamma_f = \frac{\beta\theta}{\theta+\omega-\omega\theta+\omega\beta\theta}\,, \quad \gamma_b = \frac{\omega}{\theta+\omega-\omega\theta+\omega\beta\theta}\,, \tag{2.4}$$

$E_t\pi_{t+1}$ is expected inflation at time $t$, $s_t$ represents real marginal costs (expressed as a percentage deviation with respect to its steady-state value) and $\upsilon_t$ is unexpected inflation. The parameter $\gamma_f$ determines the forward-looking component of inflation and $\gamma_b$ its backward-looking part; $\beta$ is the subjective discount rate.

Gali and Gertler rewrite the above NKPC model in terms of orthogonality conditions. Two different normalizations are used for this purpose.[5] The first one [orthogonality specification (1)] is given by

$$E_t\{[\phi\pi_t - (1-\omega)(1-\theta)(1-\beta\theta)s_t - \beta\theta\pi_{t+1} - \omega\pi_{t-1}]z_t\} = 0 \tag{2.5}$$

where $\phi = (\theta+\omega-\omega\theta+\omega\beta\theta)$, while the second one [orthogonality specification (2)] is

$$E_t[\left(\pi_t - \lambda s_t - \gamma_f\pi_{t+1} - \gamma_b\pi_{t-1}\right) z_t] = 0\,. \tag{2.6}$$

The vector $z_t$ includes variables that are orthogonal to $\upsilon_{t+1}$, allowing for GMM estimation. Quarterly U.S. data are used, with $\pi_t$ measured by the percentage change in the GDP deflator, and real marginal costs given by the logarithm of the labour income share. The instruments include four lags of inflation, labour share, commodity-price inflation, wage inflation, the long-short interest rate spread, and output gap (measured by a detrended log GDP).

Gali and Gertler's estimations yield the following values for $(\omega,\, \theta,\, \beta)$ : $(0.27,\, 0.81,\, 0.89)$ for specification (1), and $(0.49,\, 0.83,\, 0.91)$ for specification (2). When the subjective discount rate is restricted to one, the estimates are $(0.24,\, 0.80,\, 1.00)$ and $(0.52,\, 0.84,\, 1.00)$, respectively. The implied slopes are all positive and deemed to be statistically significant using IV-based asymptotic standard errors, and the fact that the overidentifying restrictions are not rejected by the $J$ test. Accordingly, Gali and Gertler conclude that there is good empirical support for the NKPC. Furthermore, the forward-looking component of inflation is more important than the backward-looking part (*i.e.* the estimated value of $\gamma_f$ is larger than the one for $\gamma_b$).

However, given the severity of the size distortions induced by weak instruments, it is important to ascertain that these results are not invalidated by such problems.[6] Ma (2002) uses corrected GMM inference methods developed by Stock and Wright (2000) to reevaluate the empirical relevance

[5] In Gali and Gertler (1999), the orthogonality conditions are written for the case $\omega = 0$; see Gali et al. (2001) for the general case.

[6] For a detailed discussion on weak instruments and their effects (as discussed in the introduction) see Nelson and Startz (1990*a*, 1990*b*), Buse (1992), Choi and Phillips (1992), Maddala and Jeong (1992), Angrist and Krueger (1994), McManus, Nankervis and Savin (1994), Bound, Jaeger and Baker (1995), Cragg and Donald (1996), Hall, Rudebusch and Wilcox (1996), Dufour (1997), Staiger and Stock (1997), Wang and Zivot (1998), Zivot et al. (1998), Stock and Wright (2000), Dufour and Jasiak (2001), Hahn and Hausman (2002, 2003), Kleibergen (2002), Moreira (2003*a*, 2003*b*), Stock et al. (2002), Kleibergen and Zivot (2003), Wright (2003); several additional papers are also cited in Dufour (2003) and Stock et al. (2002).

of the NKPC specifications. The corrected 90 per cent confidence sets (called $S$-sets) that Ma calculates are very large, including all parameter values in the interval $[0,\ 3]$ for two of the structural parameters, and $[0,\ 8]$ for the third one. Since all parameter combinations derived from these value ranges are compatible with the model, this suggests that parameters are weakly identified. We will now reassess the NKPC model using identification-robust (or weak-instrument robust) methods.

## 3. Statistical framework and methodology

We consider here two econometric specifications in order to assess Gali and Gertler's NKPC. These are given by:

$$\pi_t = \lambda s_t + \gamma_f \pi_{t+1} + \gamma_b \pi_{t-1} + u_{t+1}\,, \quad t = 1,\,\ldots\,,T\,, \tag{3.1}$$

and

$$\pi_t = \lambda s_t + \gamma_f \tilde{\pi}_{t+1} + \gamma_b \pi_{t-1} + u_{t+1}\,, \quad t = 1,\,\ldots\,,T\,, \tag{3.2}$$

where $\tilde{\pi}_t$ is a survey measure of inflation expectations. These two models differ by their assumptions on the formation of inflation expectations. In (3.1), expected inflation $E_t \pi_{t+1}$ is proxied by the realized value $\pi_{t+1}$, while in (3.2) it is replaced by the survey-based measure $\tilde{\pi}_{t+1}$ of expected inflation for $\pi_{t+1}$. It is easy to see that both approaches raise error-in-variable problems and the possibility of correlation between explanatory variables and the disturbance term in the two above equations. Studies such as Roberts (1997, 1998, 2001) have noted that the maintained specification for how expectations are formed have important implications for the empirical validity of the curve. That is, additional lags not implied by the NKPC under rational expectations may be required, even if the model is conditional on labour costs.

The parameters $\lambda$, $\gamma_f$, and $\gamma_b$, defined in equations (2.3), are nonlinear transformations of the "deep parameters" $\omega$, $\beta$, and $\theta$. The statistical details underlying our inference methodology are presented in Appendix B, where to simplify presentation, we adopt the following notation: $y$ is the $T$-dimensional vector of observations on $\pi_t$, $Y$ is the $T \times 2$ matrix of observations on $s_t$ and either of $\pi_{t+1}$ and $\tilde{\pi}_{t+1}$, $X_1$ is the vector of observations on the inflation lag $\pi_{t-1}$, $X_2$ is the $T \times k_2$ matrix of the instruments (we use 24 instruments, see section 4) and $u$ is the $T$-dimensional vector of error terms $u_t$.

The methodology we consider can be summarized as follows. To obtain a confidence set with level $1 - \alpha$ for the deep parameters, we invert the $F$-test presented in Appendix B associated with the null hypothesis

$$H_0 : \omega = \omega_0,\ \beta = \beta_0,\ \theta = \theta_0 \tag{3.3}$$

where $\omega_0, \beta_0$, and $\theta_0$ are known values. Formally, this implies collecting the values $\omega_0$, $\beta_0$, and $\theta_0$ that are not rejected by the test (i.e. the values for which the test is not significant at level $\alpha$). Taking equation (3.2) as an example, the test under consideration proceeds as follows (further discussion and references are provided in Appendix B).

1. Solve (2.3)-(2.4) for the values of $\lambda$, $\gamma_f$ and $\gamma_b$ associated with $\omega_0$, $\beta_0$, and $\theta_0$ : we denote these by $\lambda_0$, $\gamma_{f0}$ and $\gamma_{b0}$.

2. Consider the regression [which we will denote the AR-regression, in reference to Anderson and Rubin (1949)] of

$$\left\{\pi_t - \lambda_0 s_t - \gamma_{f0}\tilde{\pi}_t - \gamma_{b0}\pi_{t-1}\right\} \text{ on } \{\pi_{t-1} \text{ and the } \textit{instruments}\}. \tag{3.4}$$

   Under the null hypothesis [specifically (3.2)-(3.3)], the coefficients of the latter regression should be zero. Hence testing for a zero null hypothesis on all response coefficients in (3.4) provides a test of (3.3).

3. Compute the standard $F$-statistic for the exclusion of all regressors, namely,

$$\{\pi_{t-1} \text{ and the } \textit{instruments}\}$$

   in the regression (3.4) [see (B.13) in Appendix B]. In this context, the usual classical regression framework applies, so the latter $F$-statistic can be compared referred to its usual $F$ or $\chi^2$ cut-off points.

Tests of this type were originally proposed by Anderson and Rubin (1949) for linear Gaussian simultaneous equations models. The AR approach transforms a structural equation such as (3.2) into the regular regression framework as in (3.4), for which standard finite-sample and asymptotic distributional theory applies. The required transformation is extremely simple, despite the complexity of the model under test. Indeed, the basic test we use for inference on $\omega_0$, $\beta_0$, and $\theta_0$ differs from a standard IV-based Wald or $t$-type one by the fact that it avoids directly estimating the structural equation in (3.2), which faces identification difficulties. In contrast, the AR-regression (3.4) satisfies the usual classical regression assumptions (because no "endogenous" variables appear on its right-hand side). Whereas any statistical analysis of (3.2) requires identification constraints, these are no longer needed for inference on the regression (3.4). As shown more rigorously in Appendix B, the AR-regression provides information on the structural parameters because it is linked to the reduced form associated with the structural equation (3.2). By identification-robust, we mean here that the $F$-test is valid whether the model is identified or not.[7]

Transforming the test problem to the AR-regression framework however comes at some cost: the identification-robust $F$-test requires assessing [in regression (3.4)] the exclusion of $\pi_{t-1}$ and the 24 available instruments (25 constraints), even though the number of structural parameters under test is only 3. Instrument abundance thus leads to degrees-of-freedom losses with obvious consequences on test power. It is possible to characterize what an “optimal” instrument set looks like from the viewpoint of maximizing test power: up to a nonsingular transformation, the latter (say $\bar{Z}$) should be the mean of the endogenous explanatory variables in the model or, which is equivalent,

$$X_2 \times \{\text{the coefficient of } X_2 \text{ in the first stage regression, assumed known}\},$$

[7]We emphasize in Appendix B that the latter test will be exactly size-correct exactly if we can strictly condition on the regressors and particularly the instruments for statistical analysis; weakly exogenous regressors in our dynamic model with instruments orthogonal to the regression error terms are not in accord with the latter assumption. Nevertheless, the tests are still identification-robust. An exact test can still be devised for the NKPC model at hand despite its dynamic econometric specification if one is willing to consider strongly exogenous instruments.

where $X_2$ (as defined above) refers to the matrix of available *instruments*; see Dufour and Taamouti (2003*b*) and Appendix B of this paper. Here, the first stage regression is the regression of the left-hand side endogenous variables in (3.2) [marginal cost and expected inflation] on the included exogenous variable [the inflation lag] and $X_2$. More precisely, this involves applying steps 1-3 above after replacing the *instruments* by $\bar{Z}$, whose dimension is $T \times 2$. So, the optimal identification-robust $F$-test requires assessing [in the regression (3.4)] the exclusion of $\pi_{t-1}$ and the two optimal instruments (3 constraints); recall that the number of structural parameters under test is indeed 3. This provides optimal information reduction, which improves the power of the test (and thereby may tighten the confidence sets based on these tests).

In practice, however, the coefficient of $X_2$ in the first-stage regression ($\Pi_2$ in Appendix B) is not known, and estimates of this parameter must be "plugged in", which of course only leads to an “approximately optimal” procedure. As described in Dufour (2003), many procedures that aim at being identification-robust as well as improving the AR procedure from the viewpoint of power rely on different choices of $\bar{Z}$. In particular, if a constrained OLS estimator imposing the structure underlying (3.2) is used [$\hat{\Pi}_2^0$ in equation (B.15)], then the associated procedure yields Kleibergen’s (2002) K-test.[8] In other words, Kleibergen’s (2002) test obtains on applying steps 1-3 above, replacing the *instruments* by

$$\bar{Z}_K = X_2 \hat{\Pi}_2^0 \,.$$

To avoid confusion, the tests based on $X_2$ and $\bar{Z}_K$ are denoted by AR and AR-K, respectively. This is the alternative “parsimonious identification-robust” procedure we shall consider here.

Finally, we invert these tests to get confidence sets as follows: using a grid search over the economically meaningful set of values for $\omega$, $\beta$, and $\theta$, we sweep the economically relevant choices for $\omega_0$, $\beta_0$, and $\theta_0$.[9] For each parameter combination considered, we compute the statistics AR and AR-K as described above and their respective $p$-values. The parameter vectors for which the $p$-values are greater than the level $\alpha$ constitute a confidence set with level $1-\alpha$. Since every choice of $\omega_0$, $\beta_0$, and $\theta_0$ entails [using (2.3)] a choice for $\lambda$, $\gamma_f$ and $\gamma_b$, this procedure also yields conformable confidence sets for the latter parameters. These confidence sets reflect the structure, and obtain without further computations, although $\lambda$, $\gamma_f$ and $\gamma_b$ are transformations of the deep parameters. Therein lies a significant advantage in using our approach as an alternative to standard nonlinear Wald-based techniques.

To conclude, it is worth to emphasize two points. First, if the confidence set obtained by inverting an AR-type test is empty, *i.e.* if no economically acceptable value of the model deep parameters is upheld by the data, then we can infer that the model is rejected at the chosen significance level. We thus see that the procedure used here may be seen as an identification-robust alternative to the standard GMM-based $J$ test. In the same vein, utterly uninformative (too wide) confidence

[8] To correct for plug-in estimation effects (*i.e.* for estimating $\Pi_2$), Dufour and Jasiak (2001), Dufour and Taamouti (2003*b*, 2003*a*) recommend split sample estimation techniques, where the first sub-sample is used to estimate $\Pi_2$ and the second sub-sample is used to run the optimal AR-test based on the latter estimate. Results applying these versions of the tests are available from the authors upon request.

[9] We allow the range (0, 1) as the admissible space for each of $\omega$, $\theta$, and $\beta$. The values are varied with increments of 0.03 for $\omega$ and $\theta$, and by 0.01 for $\beta$. The increment of 0.03 was chosen for the first two parameters (rather than 0.01) to minimize the computational burden.

sets allow one to assess model fit, since unbounded confidence sets do occur under identification difficulties [see the discussion in Dufour (2003)]. Our procedure (which performs, for practical purposes, the same specification checks conveyed as $J$-type test) has a clear "built-in" advantage over GMM-based $t$-type confidence intervals, backed by a non-significant $J$ test.[10]

Our procedure offers another important advantage not shared by the latter standard approach. So far, we have considered the estimation and test problem given a specific significance (or confidence) level $\alpha$. Alternatively, the $p$-value associated with the above defined tests, which provides a formal specification check, can be used to assess the empirical fit of the model. In other words, the values (uniqueness is not granted) of $\omega_0$, $\beta_0$, and $\theta_0$ that lead to the largest $p$-value formally yield the set of "least rejected" models, *i.e.* models that are most compatible with the data.[11] In practice, analyzing the economic information content of these least rejected models (associated with the least rejected "deep parameter" combinations) provides decisive and very useful goodness-of-fit checks.

# 4. Empirical results

We applied the above-defined inference methods to the hybrid NKPC models in (3.1) and (3.2) for both U.S. and Canadian data. One difference between our specifications and those of Gali and Gertler is that we use a real-time output-gap measure in the set of instruments instead of a gap detrended using the full sample. The latter measure does not appear to be an appropriate instrument since, when the full sample is used, lagged values of the gap are, by construction, related to future information. To avoid this, we proceed iteratively: to obtain the value of the gap at time $t$, we detrend GDP with data ending in $t$. The sample is then extended by one more observation and the trend is reestimated. This is used to detrend GDP and yields a value for the gap at time $t + 1$. This process is repeated until the end of the sample. In this fashion, the gap measure at time $t$ does not use information beyond that period and can therefore be used as a valid instrument. We also considered a quadratic trend for this purpose.[12]

Regarding survey expectations, the Federal Reserve Bank of Philadelphia publishes quarterly mean forecasts of the next quarter's U.S. GDP implicit price deflator. We first-difference this series to obtain our inflation-expectations series for the U.S.[13] In the case of Canada, the survey-based inflation expectations series were obtained from Canada's Conference Board Survey; further details on the Canadian data appear in Appendix A. For the remaining variables, other than the output gap, we use the Gali and Gertler data and instrument set for the U.S., and the corresponding variables in the case of Canada. Because of the expectations variables in the data set, our samples start in 1970Q1. The data end in 1997Q4 for the U.S., and in 2000Q4 for Canada.

[10] Indeed, if the AR confidence set with level $1 - \alpha$ is empty, then the usual LIML over-identification test statistic will exceed a specific bounds-based identification-robust $\alpha$-level critical point, *i.e.* the associated over-identification test is conclusively significant at level $\alpha$.

[11] This method underlies the principles of Hodges-Lehmann estimators; see Hodges and Lehmann (1963, 1983). Least-rejected values may thus be interpreted as "point estimates".

[12] We repeated our estimations using a cubically-detrended real-time gap measure, as well the Christiano-Fitzgerald one-sided band-pass filter, and obtained qualitatively similar results.

[13] Source:http://www.phil.frb.org/econ/spf/index.html.

Table 1. Anderson-Rubin tests with rational expectations

| Test Type | | Unrestricted model | | | |
|---|---|---|---|---|---|
| | | Max $p$-value | Deep parameters $(\omega, \theta, \beta)$ | Reduced-form parameters $(\lambda, \gamma_f, \gamma_b)$ | Freq. |
| AR | U.S. | 0.2771 | (0.40, 0.64, 0.96) | (0.08, 0.60, 0.39) | 2.78 |
| | Canada | - | - | - | - |
| AR-K | U.S. | 0.9993 | (0.40, 0.61, 0.98) | (0.09, 0.59, 0.40) | 2.56 |
| | Canada | 0.9990 | (0.01, 0.37, 0.21) | (1.53, 0.21, 0.03) | 1.59 |
| | | $\beta = 0.99$ | | | |
| | | Max $p$-value | Deep Parameters $(\omega, \theta, \beta)$ | Reduced-form Parameters $(\lambda, \gamma_f, \gamma_b)$ | Freq. |
| AR | U.S. | 0.2765 | (0.37, 0.64, 0.99) | (0.08, 0.63, 0.37) | 2.78 |
| | Canada | - | - | - | - |
| AR-K | U.S. | 0.9987 | (0.37, 0.64, 0.99) | (0.08, 0.63, 0.37) | 2.78 |
| | Canada | 0.2900 | (0.01, 0.10, 0.99) | (7.30, 0.91, 0.09) | 1.11 |

Note - AR is the Anderson-Rubin test and AR-K refers to the Kleibergen test. Freq. is the average frequency of price adjustment, measured in quarters.

We first apply the AR test to the U.S. data, and for equation (3.1), to assess the Gali and Gertler (1999) reported estimates. Specifically, we test whether $\omega$, $\theta$, and $\beta$ are $(0.27,\ 0.81,\ 0.89)$ or (0.49, 0.83, 0.91), which correspond to those authors' estimates based on their orthogonality specifications (1) and (2), respectively. We find all tests to be significant at conventional levels, so that their estimated parameter values are rejected. We then ask whether, for the same instrument set, there exists a value of the parameter vector for which the hybrid NKPC is not rejected. Interestingly, we find some dramatically different results depending on whether (3.1) or (3.2) is used.

For the U.S. rational expectation solution, we find a bounded but fairly large confidence set. This entails that there is a multitude of different parameter combinations which are compatible with the econometric model tested, although the set is much smaller than the S-sets constructed by Ma.[14] However, for the model using survey expectations the confidence set is empty (at the 95% level). Thus, not a single parameter value combination is compatible with this particular econometric model, implying that with survey expectations, the model is not appropriate. Regarding the Canadian data, we find that the outcomes are reversed. Thus, it is the model with rational expec-

[14]There is a slight difference between our two instrument sets: Ma's set includes a constant and has no fourth lag for any of the variables in levels.

Table 2. Anderson-Rubin tests with survey expectations

| Test type | | Unrestricted model | | | |
|---|---|---|---|---|---|
| | | Max $p$-value | Deep parameters $(\omega,\, \theta,\, \beta)$ | Reduced-form parameters $(\lambda,\, \gamma_f,\, \gamma_b)$ | Freq. |
| AR | U.S. | - | - | - | - |
| | Canada | 0.1009 | $(0.01, 0.97, 0.89)$ | $(0.00, 0.88, 0.01)$ | 33.33 |
| AR-K | U.S. | 0.9983 | $(0.01, 0.61, 0.64)$ | $(0.38, 0.63, 0.02)$ | 2.56 |
| | Canada | 0.0890 | $(0.01, 0.97, 0.90)$ | $(0.00, 0.89, 0.01)$ | 33.33 |
| | | $\beta = 0.99$ | | | |
| | | Max $p$-value | Deep parameters $(\omega,\, \theta,\, \beta)$ | Reduced-form parameters $(\lambda,\, \gamma_f,\, \gamma_b)$ | Freq. |
| AR | U.S. | - | - | - | - |
| | Canada | 0.0562 | $(0.01, 0.97, 0.99)$ | $(0.00, 0.98, 0.01)$ | 33.33 |
| AR-K | U.S. | 0.6057 | $(0.52, 0.22, 0.99)$ | $(0.40, 0.29, 0.70)$ | 1.28 |
| | Canada | - | - | - | - |

Note - AR is the Anderson-Rubin test and AR-K refers to the Kleibergen test. Freq. is the average frequency of price adjustment, measured in quarters.

tations that generates the empty confidence set, while the specification using survey data yields the non-empty one. The latter is so small that there are only some parameter value combinations for which the model is statistically valid.

Along with the identification-robust confidence sets, one of the great advantages of using the Anderson-Rubin method is that it yields the parameter combination that is least rejected, or, alternatively, that has the highest $p$-value. Formally, as explained in the previous section, this point estimate corresponds to the so-called Hodges-Lehmann estimate and can be compared with point estimates obtained using more conventional estimation methods (such as GMM). We report this estimate for the U.S. and Canada in the upper panels of Tables 1 and 2, respectively. From here, we can see that, under rational expectations, the values of the deep parameters $(\omega,\, \theta,\, \beta)$ that correspond to the maximal $p$-value for the U.S. is given by $(0.40,\, 0.64,\, 0.96)$. These translate into a value of 0.6 for the coefficient of the forward-looking component on inflation $(\gamma_f)$, and 0.39 for the coefficient of the backward-looking component $(\gamma_b)$. Furthermore, the coefficient on the marginal cost variable is 0.08, and the average frequency of price adjustment is 2.78 quarters.

Based on the Hodges-Lehmann estimates, the findings provide support for the optimization-based Phillips curve, and the notion that the forward-looking component of the U.S. inflation process is more important than its backward-looking part. In addition, the estimate for the average frequency

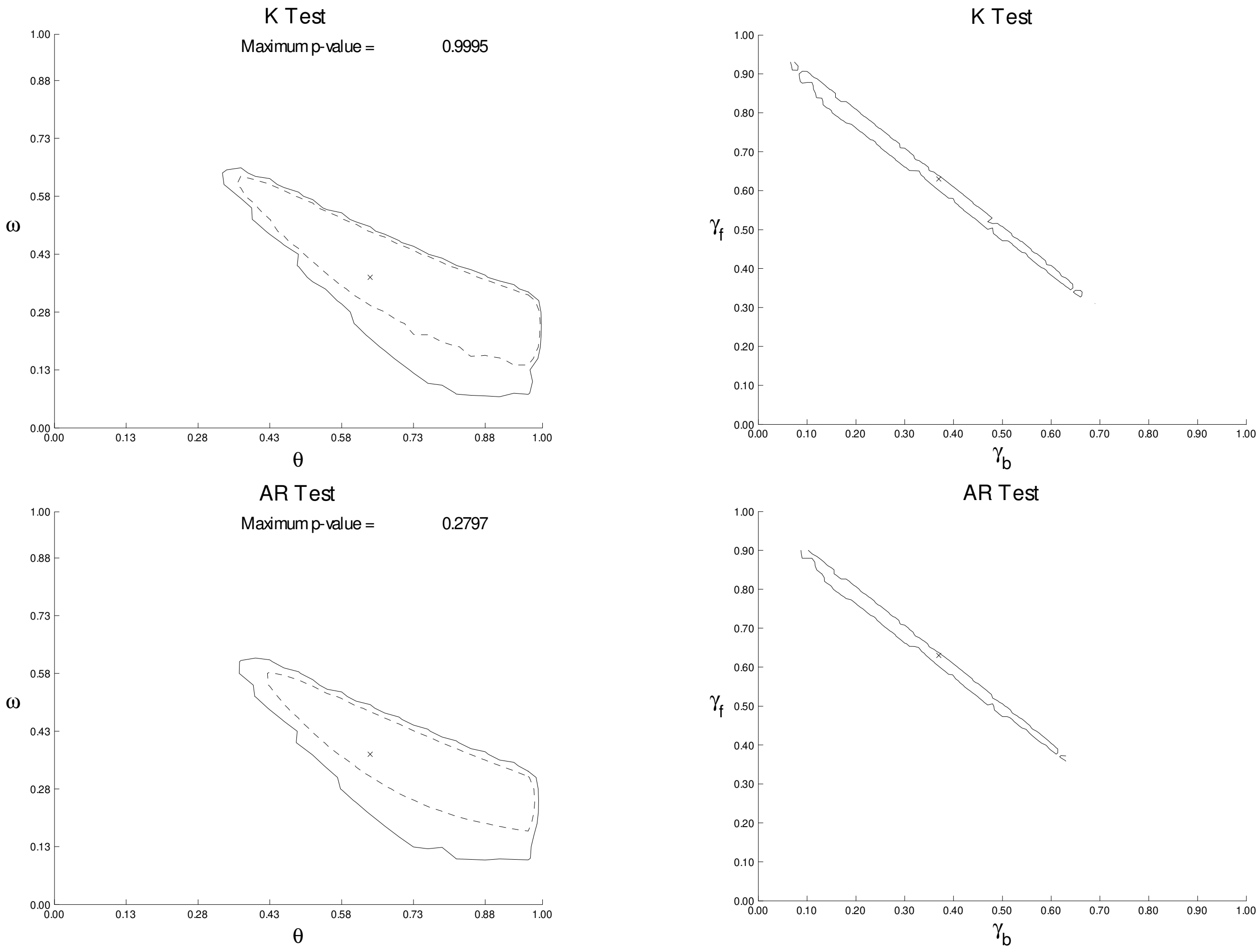
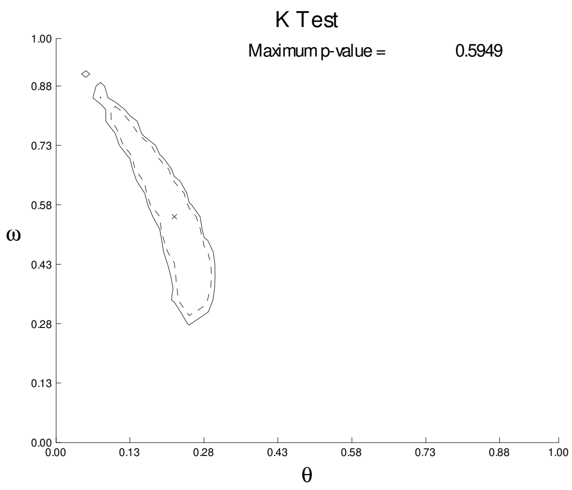



Figure 1. AR and AR-K tests (U.S., Rational Expectations).



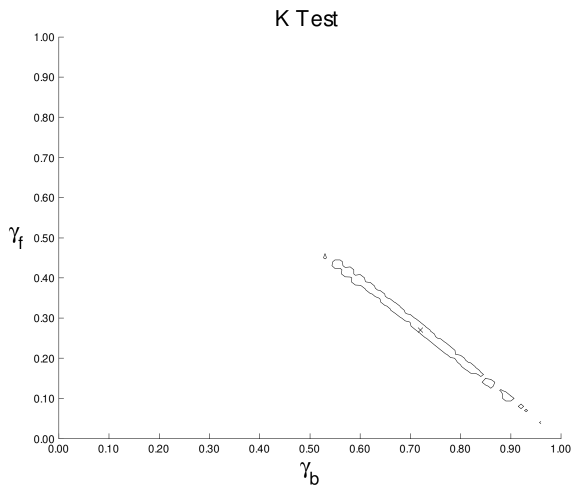




Figure 2. AR-K tests (U.S., Survey Expectations).

of price adjustment is fairly close to the value of 1.8 obtained based on micro data [see, for example, Bils and Klenow (2004)].[15] On the other hand, the graphs in the lower panel of Figure 1 provide a qualification to the above statement.

The graph in the bottom panel, on the left, depicts the $95\%$ (solid line, $p$-value $= 0.05$) and $90\%$ (dashed line, $p$-value $= 0.10$) confidence sets based on the AR test, and for the case where the subjective discount parameter is constrained to lie between 0.95 and 0.99. An "X" marks the spot corresponding to the highest $p$-value obtained (0.2797). Immediately, three features can be noticed: (i) the sets of parameter values that the test does not reject at the 5 and 10% levels are fairly large, (ii) within these sets, there is more than one $\omega$ value that corresponds to a given $\theta$, and vice-versa, and (iii) the parameter combination that yields the highest $p$-value is very close to points that have a $p$-value of 0.10 only. In other words, even when $\beta$ is constrained quite tightly, the uncertainty regarding the estimated values of the other parameters is relatively high. This is seen more easily in the adjacent graph which depicts the values corresponding to the $95\%$ confidence set in the $\gamma_f$ and $\gamma_b$ space. Notice, in particular, that a value of 0.60 for the backward-looking component of inflation, and 0.37 for the forward-looking part is as likely to be obtained as a value of 0.90 and 0.10 for the forward and backward-looking components, respectively.

Turning now to Canadian data, recall that the model with rational expectations is not compatible with the data, but that the one with survey expectations does yield a non-empty set. The results corresponding to the highest $p$-value for the latter are reported in Table 2. In this case, the maximal $p$-value is 0.1009 while the deep parameters are $(0.01,\ 0.97, 0.89)$. Based on the fact that the proportion of firms that follow a rule-of-thumb is practically zero ($\omega = 0.01$), we would conclude that a purely forward-looking model is applicable to Canada. However, a look at the reduced-form parameters and the average frequency of price adjustment indicate that the model is economically not plausible. This is the case even if $\beta$ is constrained to 0.99 in the estimation.[16]

Results based on Kleibergen's statistic are also reported in Tables 1 and 2. As for the AR tests, two sets of outcomes are tabulated for each country: the parameter values that yield the highest test $p$-value for the unrestricted model appear in the upper panel, while the lower one shows the corresponding elements when $\beta$ is constrained to 0.99.

Let us first examine the results for the U.S. with the rational expectations model. When $\bar{Z}_K$ is used as the instrument set, the model is least rejected for the parameter combination (0.40, 0.61, 0.98), and the $p$-value is 0.9993. These values are extremely close to those reported for the corresponding restricted estimation (with $\beta$ constrained to 0.99) case, and also, to those of the AR tests.

With the model based on survey expectations (Table 2), although the AR test yields an empty confidence set for the U.S., the test that corresponds to Kleibergen's K-test (the AR-K test) yields a least-rejected parameter combination that suggests strongly forward-looking behaviour ($\gamma_f = 0.63$, $\gamma_b = 0.02$). In addition, when the subjective discount rate is constrained to 0.99, the AR-K test now points to a much more important backward-looking component for inflation.

Our findings are somewhat similar with Canadian data. Although the AR-K test yields outcomes

[15] Gali and Gertler report average price adjustment frequencies of roughly 5 quarters.

[16] For this reason, and because all of the admissible $\omega$ values in the AR-based confidence sets equal 0.01, no figures are provided for Canada.

similar to those of the AR test for the unrestricted model with survey expectations, with rational expectations, the AR-K yields parameter values that suggest a less important forward-looking role in inflation. In addition, the estimate for the average frequency of price adjustment is 1.6, very much in line with micro data [as in Bils and Klenow (2004)]. These results are nevertheless difficult to reconcile with the value for $\lambda$, which is essentially zero. In addition, once the subjective discount rate parameter is constrained to 0.99, the conclusions on the rational expectations specification from the AR-K test point to a much more important forward-looking component of inflation ($\gamma_f = 0.91$, $\gamma_b = 0.09$). The unusual feature in this case is the value of the coefficient on the marginal cost variable, $\lambda$, which stands at 7.30.

Figures 1 and 2 present U.S. results for the AR-K test for the case where $\beta$ is constrained to fall between 0.95 and 0.99. Under rational expectations (Figure 1), the confidence set based on inverting the AR-K test is larger than that based on the AR but results are in line with each other, in the sense that the 95% confidence sets are more skewed towards higher $\gamma_f$ than $\gamma_b$. Turning to Figure 2, we find that the AR-K test produces strong support for a larger backward-looking component to inflation.

Taken collectively, the results in this section point to problems of weak identification in these models. Nevertheless, we find that there is some support for the hybrid NKPC for the U.S., whereas the model is not suited to Canada.

## 5. Conclusion

In this paper we used finite-sample methods to test the empirical relevance of Gali and Gertler's (1999) NKPC equations, using AR tests as well as Kleibergen's more parsimonious procedure. We focused on the Gali and Gertler's (1999) specification, for both U.S. and Canadian data. Two variants of the model were studied: one based on a rational-expectations assumption, and a modification to the latter which consists in using survey data on inflation expectations. In the U.S. case, Gali and Gertler's (1999) original data set were used except for the output gap measure and survey expectations where applicable.

First, we found some evidence of identification difficulty. Nevertheless, the maximal $p$-value arguments point out those parameter values for which the model is least rejected – a very useful feature of our proposed identification-robust techniques. Second, we found support for Gali and Gertler's hybrid NKPC specification with rational expectations for the U.S. Third, neither model was found to be well suited to describe inflation dynamics in Canada. Fourth, we found that, for the cases where the Anderson-Rubin test yields an empty confidence set, the AR-K procedure leads to conflicting results for the restricted and unrestricted models.

These results underscore the need for employing identification-robust inference in the estimation of expectations-based dynamic macroeconomic relations.

# Appendix

## A. Data description for Canada

The inflation expectations series is obtained from the Conference Board of Canada survey. Each quarter, participants are asked to forecast the annual average (GDP-deflator) inflation rate for the current year. Let us denote $\tilde{\pi}_1^a$, $\tilde{\pi}_2^a$, $\tilde{\pi}_3^a$, and $\tilde{\pi}_4^a$, the annual average inflation forecasts made in quarters 1, 2, 3, and 4 of a given year, respectively. Clearly, forecasts that are made in the second, third, and fourth quarters are likely to integrate realized (and observed) inflation in quarters 1, 1 and 2, and 1, 2 and 3, respectively.

To obtain a "pure" quarterly expectations series, we proceed as follows: First, denote the forecasted quarterly inflation rate in quarters 1 to 4 as $\tilde{\pi}_1^q$, $\tilde{\pi}_2^q$, $\tilde{\pi}_3^q$, and $\tilde{\pi}_4^q$, respectively. Similarly, let $\pi_1^q$, $\pi_2^q$, $\pi_3^q$ be the realized quarterly inflation rates in quarters 1, 2, and 3, respectively. Then, the forecasted quarterly inflation rates are calculated as follows:

$$
\begin{aligned}
\tilde{\pi}_1^q &= \tilde{\pi}_1^a/4 \\
\tilde{\pi}_2^q &= (\tilde{\pi}_2^a - \pi_1^q)/3 \\
\tilde{\pi}_3^q &= (\tilde{\pi}_3^a - \pi_1^q - \pi_2^q)/2 \\
\tilde{\pi}_4^q &= (\tilde{\pi}_4^a - \pi_1^q - \pi_2^q - \pi_3^q)
\end{aligned}
$$

The remaining data are quarterly time series from Statistics Canada's database.
Any monthly data are converted to quarterly frequency.

**Output gap** is the deviation of real GDP ($y_t = \ln Y_t$) from its steady state, approximated by a quadratic trend: $\hat{y} = 100(y_t - \bar{y}_t)$, where $Y_t = I56001 - I56013 - I56018$.

**Price inflation** is the quarterly growth rate of the total GDP deflator:
$\pi_t = 100(\ln P_t - \ln P_{t-1})$ and $P_t = D15612$

**Wage inflation** is the quarterly growth rate of compensation of employees:
$w_t = 100(\ln W_t - \ln W_{t-1})$, where $W_t = D17023/N_t$.
$N_t = LFSA201$ for 1970:1-1975:4 and $N_t = D980595$ for 1976:1-2000:4

**Labour income share** is the ratio of total compensation and nominal GDP: $ls_t = \ln S_t$, and $s_t = 100(ls_t - s)$, the labour income share in deviation from its steady-state, where $s = \ln S, S = \sum_t^T \ln(S_t)/T$ and $S_t = (D17023 - D17001)/(D15612 * Y_t)$.

**Average real marginal costs for a Cobb-Douglas function**: $rmc_t^{avg} = s_t$.

## B. The AR test and related procedures

Consider the following structural equation:

$$y = Y\delta + X_1\kappa + u, \tag{B.1}$$

where $y$ is a $T \times 1$ dependent variable, $Y$ is a $T \times m$ matrix of endogenous variables, $X_1$ is a $T \times k_1$ matrix of exogenous variables, and $u$ is an error term that satisfies standard regularity conditions typical of IV regressions; see Dufour and Jasiak (2001). In our context (see section 3), $y$ is the $T$-dimensional vector of observations on $\pi_t$, $Y$ is the $T \times 2$ matrix of observations on $s_t$ and $\tilde{\pi}_{t+1}$ [or $\pi_{t+1}$, depending on the context], $X_1$ is the vector of observations on the inflation lag $\pi_{t-1}$, $X_2$ is the $T \times k_2$ matrix of the instruments, and $u$ is the $T$-dimensional vector of error terms $u_t$.

Suppose that the reduced form associated with the right-hand-side endogenous regressors is

$$Y = X_1\mathit{\Pi}_1 + X_2\mathit{\Pi}_2 + V \tag{B.2}$$

where $V$ is an $T \times m$ matrix of error terms assumed to be cross-correlated and correlated with $u$, and $X_2$ is the matrix of available instruments.[17] In this case, the reduced form associated with (B.1) is

$$\begin{aligned} y &= X_1 p_1 + X_2 p_2 + u + V\delta, && \text{(B.3)} \\ p_1 &= \mathit{\Pi}_1\delta + \kappa, \qquad p_2 = \mathit{\Pi}_2\delta. && \text{(B.4)} \end{aligned}$$

Identification constraints follow from (B.4) and amount to the rank condition

$$rank(\mathit{\Pi}_2) = m. \tag{B.5}$$

Consider hypotheses of the form

$$H_0 : \delta = \delta_0. \tag{B.6}$$

In this case, the transformed model

$$y - Y\delta_0 = Y(\delta - \delta_0) + X_1\kappa + u,$$

has reduced form

$$y - Y\delta_0 = X_1[\mathit{\Pi}_1 (\delta - \delta_0) + \kappa] + X_2[\mathit{\Pi}_2 (\delta - \delta_0)] + u + V (\delta - \delta_0). \tag{B.7}$$

In view of this, the AR test assesses the exclusion of $X_2$ (of size $T \times k_2$) in the regression of $y - Y\delta_0$

[17] In Dufour and Taamouti (2006) and Dufour (2003), we stress that: (i) linearity of the latter reduced form is strictly not necessary, (ii) further exogenous regressors ("excluded" instruments) may enter into the equation in addition to the instrument set. To present the test in its simplest form, we maintain the standard linear form (B.2) and refer the reader to later references for disucussion of the more general setting. Note that the assumptions regarding the reduced form for $Y$ do not affect the actual implementation of the test, so our simplified presentation does not lack generality for practical purposes.

on $X_1$ and $X_2$, which can be conducted using a standard $F$-test. Let $X = (X_1, X_2)$, and define

$$M = I - X(X'X)^{-1}X', \qquad M_1 = I - X_1(X_1'X_1)^{-1}X_1'.$$

The statistic then takes the form

$$AR(\delta_0) = \frac{(y - Y\delta_0)'(M_1 - M)(y - Y\delta_0)/k_2}{(y - Y\delta_0)' M (y - Y\delta_0)/(T - k_1 - k_2)}. \tag{B.8}$$

Under the null hypothesis, assuming strong exogeneity and identically, independently distributed (i.i.d.) normal errors,

$$AR(\delta_0) \sim F(k_2, T - k_1 - k_2). \tag{B.9}$$

Following the usual classical regression analysis, the latter strong hypotheses on the error terms can be relaxed so that, under standard regularity conditions,

$$k_2\, AR(\delta_0) \overset{asy}{\sim} \chi^2(k_2). \tag{B.10}$$

It is important to emphasize that identification constraints do $not$ intervene. In other words (B.9) or (B.10) hold whether (B.5) is verified or not; this is what "identification robustness" usually means. The test can be readily extended to accommodate additional constraints on the coefficients of (the full vector or a any subset of) the $X_1$ variables. For example, the hypothesis

$$H_0 : \delta = \delta_0,\ \kappa = \kappa_0, \tag{B.11}$$

can be assessed in the context of the transformed regression

$$\begin{aligned} y - Y\delta_0 - X_1\kappa_0 \quad &= \quad X_1[\mathit{\Pi}_1(\delta - \delta_0) + (\kappa - \kappa_0)\} \\ &\qquad + X_2[\mathit{\Pi}_2(\delta - \delta_0)] + u + V(\delta - \delta_0) \end{aligned} \tag{B.12}$$

which leads to the following $F$-statistic

$$AR(\delta_0, \kappa_0) = \frac{(y - Y\delta_0 - X_1\kappa_0)'(I - M)(y - Y\delta_0 - X_1\kappa_0)/(k_1 + k_2)}{(y - Y\delta_0 - X_1\kappa_0)' M (y - Y\delta_0 - X_1\kappa_0)/(T - k_1 - k_2)}. \tag{B.13}$$

While the test in its original form was derived for the case where the first-stage regression is linear, we re-emphasize that it is in fact robust to: (i) the specification of the model for $Y$, and (ii) excluded instruments; in other words, the test is valid regardless of whether the first-stage regression is linear, and whether the matrix $X_2$ includes all available instruments. As argued in Dufour (2003), since one is never sure that all instruments have been accounted for, the latter property is quite important. Most importantly, this test [and several variants discussed in Dufour (2003)] is the only truly pivotal statistic whose properties in finite samples are robust to the quality of instruments.

Note that exactness strictly requires that we can condition on $X$ (*i.e.* we can take $X$ as fixed for statistical analysis). This holds particularly for the instruments. In the presence of weakly exogenous regressors, the test remains identification-robust. The intuition underlying this result is the following: conducting the test via the Anderson-Rubin regressions (B.7) and (B.12) [*which constitute statistical reduced forms*] easily transforms the test problem from the IV-regression [*which requires* (B.5)] to the classical linear regression statistical framework [*which does not require* (B.5)]. This provides an attractive solution to identification difficulties, a property not shared by IV-based Wald statistics nor GMM-based J-tests.

Despite the latter desirable statistical properties, the test as presented above provides no guidance for practitioners regarding the choice of instruments. In addition, simulation studies reported in the above-cited references show that the power of AR-type tests may be affected by the number of instruments. To see this, consider the case of (B.1)-(B.6): here, the AR test requires assessing (in the regression of $y - Y\delta_0$ on $X_1$ and $X_2$) the exclusion of the $T \times k_2$ variables in $X_2$, even though the number of structural parameters under test is $m$ (the structural parameter under test $\delta$ is $m \times 1$). On recalling that identification entails $k_2 \geq m$, we see that over-identification (or alternatively, the availability of more instruments) leads to degrees-of-freedom losses with obvious implication on power. To circumvent this problem, an optimal instrument (in the sense that it yields a *point-optimal* test) is given by

$$\bar{Z} = X_2 \Pi_2$$

where $\Pi_2$ is the coefficient of $X_2$ in the first-stage regression, i.e. the regression of $Y$ on $X_1$ and $X_2$; see Dufour and Taamouti (2003*b*). Formally, this implies applying (B.9) or (B.13), replacing $X_2$ by $\bar{Z}$ (observe that $X_2$ intervene in these statistics via $M = I - X(X'X)^{-1}X'$, where $X = (X_1,\ X_2)$).

Clearly, the latter optimal instrument involves information reduction, because the associated AR-test amounts to testing for the exclusion of the $T \times m$ variables in $\bar{Z}$, which preserves available degrees-of-freedom even if the model is highly over-identified. In other words, the optimal test can reflect the informational content of all available instruments with no statistical costs.

Unfortunately, $\Pi_2$ is unknown so the approximate optimal instruments need to be estimated, with obvious implications on feasibility and exactness. As observed in Dufour (2003), the OLS estimator

$$\hat{\Pi}_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 Y \tag{B.14}$$

of $\Pi_2$ in the unrestricted reduced form multivariate regression (B.2) is used in the construction of $\bar{Z}$, then the associated statistic coincides with the LM criterion defined by Wang and Zivot (1998). In addition, the K-statistic of Kleibergen (2002) may be interpreted as based on an approximation of the optimal instrument [see Dufour and Khalaf (2003)]. In this case, $\Pi_2$ is replaced by its constrained reduced form OLS estimates imposing the structural identification condition (B.5):

$$\hat{\Pi}_2^0 = \hat{\Pi}_2 - (X_2' M_1 X_2)^{-1} X_2' M_1 \left[y - Y\delta_0\right] \frac{\left[y - Y\delta_0\right]' MY}{\left[y - Y\delta_0\right]' M \left[y - Y\delta_0\right]}\,. \tag{B.15}$$

Wang and Zivot (1998) show that the distribution of the LM statistic is bounded by the $\chi^2(k_2)$ distribution; Kleibergen (2002) shows that a $\chi^2(m)$ cut-off point is asymptotically identification-robust for the K-statistic. To obtain an $F(m,\ .)$ or $\chi^2(m)$ cut-off point for both statistics correcting for plug-in effects, split sample methods (where the first sub-sample is used to estimate $\Pi_2$ and the second to run the AR-test based on the latter estimate) may also be exploited; see Dufour and Jasiak (2001) and Dufour and Taamouti (2003*b*).

## References


Anderson, T. W. and Rubin, H. (1949), 'Estimation of the parameters of a single equation in a complete system of stochastic equations', *Annals of Mathematical Statistics* **20**, 46–63.

Angrist, J. D. and Krueger, A. B. (1994), Split sample instrumental variables, Technical Working Paper 150, N.B.E.R., Cambridge, MA.

Bils, M. and Klenow, P. (2004), 'Some evidence on the importance of sticky prices', *The Journal of Political Economy* **112**, 947–985.

Bound, J., Jaeger, D. A. and Baker, R. M. (1995), 'Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak', *Journal of the American Statistical Association* **90**, 443–450.

Buse, A. (1992), 'The bias of instrumental variables estimators', *Econometrica* **60**, 173–180.

Choi, I. and Phillips, P. C. B. (1992), 'Asymptotic and finite sample distribution theory for IV estimators and tests in partially identified structural equations', *Journal of Econometrics* **51**, 113–150.

Cragg, J. G. and Donald, S. G. (1996), Testing overidentifying restrictions in unidentified models, Discussion paper, Department of Economics, University of British Columbia and Boston University, Boston.

Dufour, J.-M. (1997), 'Some impossibility theorems in econometrics, with applications to structural and dynamic models', *Econometrica* **65**, 1365–1389.

Dufour, J.-M. (2003), 'Identification, weak instruments and statistical inference in econometrics', *Canadian Journal of Economics* **36**(4), 767–808.

Dufour, J.-M. and Jasiak, J. (2001), 'Finite sample limited information inference methods for structural equations and models with generated regressors', *International Economic Review* **42**, 815–843.

Dufour, J.-M. and Khalaf, L. (2002), 'Simulation based finite and large sample tests in multivariate regressions', *Journal of Econometrics* **111**(2), 303–322.

Dufour, J.-M. and Khalaf, L. (2003), Simulation-based finite-sample inference in simultaneous equations, Technical report, Centre interuniversitaire de recherche en analyse des organisations (CIRANO) and Centre interuniversitaire de recherche en économie quantitative (CIREQ), Université de Montréal.

Dufour, J.-M. and Taamouti, M. (2003*a*), On methods for selecting instruments, Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M. and Taamouti, M. (2003*b*), Point-optimal instruments and generalized Anderson-Rubin procedures for nonlinear models, Technical report, C.R.D.E., Université de Montréal.

Dufour, J.-M. and Taamouti, M. (2005), 'Projection-based statistical inference in linear structural models with possibly weak instruments', *Econometrica* **73**(4), 1351–1365.

Dufour, J.-M. and Taamouti, M. (2006), 'Further results on projection-based inference in IV regressions with weak, collinear or missing instruments', *Journal of Econometrics* **forthcoming**.

Eichenbaum, M. and Fisher, J. (2004), Evaluating the Calvo model of sticky prices, Technical report, Federal Reserve Bank of Chicago.

Fuhrer, J. (1997), 'The (un)importance of forward-looking behaviour in price specifications', *Journal of Money, Credit and Banking* **29**, 338–50.

Fuhrer, J. C. and Olivei, G. P. (2004), Estimating forward looking Euler equations with GMM estimators: An optimal instruments approach, Technical report, Federal Reserve Bank of Chicago.

Fuhrer, J. and Moore, G. (1995), 'Inflation persistence', *Quarterly Journal of Economics* **110**, 127–59.

Gagnon, E. and Khan, H. (2005), 'New Phillips curve under alternative production technologies for Canada, the US, and the Euro area', *European Economic Review* **49**, 1571–1602.

Gali, J. and Gertler, M. (1999), 'Inflation dynamics: A structural econometric analysis', *Journal of Monetary Economics* **44**, 195–222.

Gali, J., Gertler, M. and Lopez-Salido, J. D. (2001), 'European inflation dynamics', *European Economic Review* **45**, 1237–1270.

Hahn, J. and Hausman, J. (2002), Weak instruments: Diagnosis and cures in empirical econometrics, Technical report, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Hahn, J. and Hausman, J. (2003), IV estimation with valid and invalid instruments, Technical report, Department of Economics, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Hall, A. R., Rudebusch, G. D. and Wilcox, D. W. (1996), 'Judging instrument relevance in instrumental variables estimation', *International Economic Review* **37**, 283–298.

Hodges, Jr, J. L. and Lehmann, E. L. (1963), 'Estimates of location based on rank tests', *The Annals of Mathematical Statistics* **34**, 598–611.

Hodges, Jr, J. L. and Lehmann, E. L. (1983), Hodges-Lehmann estimators, *in* N. L. Johnson, S. Kotz and C. Read, eds, 'Encyclopedia of Statistical Sciences, Volume 3', John Wiley & Sons, New York, pp. 642–645.

Khalaf, L. and Kichian, M. (2002), Simulation-based tests of pricing-to-market, *in* E. Kontoghiorghes, B. Rustem and S. Siokos, eds, 'Computational Methods in Decision-Making, Economics and Finance', Kluwer Academic Publishers, The Netherlands, pp. 583–603.

Khalaf, L. and Kichian, M. (2005), ‘Exact tests of the stability of the Phillips curve: The Canadian case.’, *Computational Statistics and Data Analysis* **49**, 445–460.

Kleibergen, F. (2002), ‘Pivotal statistics for testing structural parameters in instrumental variables regression’, *Econometrica* **70**(5), 1781–1803.

Kleibergen, F. and Zivot, E. (2003), ‘Bayesian and classical approaches to instrumental variable regression’, *Journal of Econometrics* **114**(1), 29–72.

Linde, J. (2001), Estimating new-Keynesian Phillips curves: A full information maximum likelihood approach, Technical report, Sveriges Riksbank Working Paper No. 129, Stockholm, Sweden.

Ma, A. (2002), ‘GMM estimation of the new Phillips curve’, *Economic Letters* **76**, 411–417.

Maddala, G. S. and Jeong, J. (1992), ‘On the exact small sample distribution of the instrumental variable estimator’, *Econometrica* **60**, 181–183.

McManus, D. A., Nankervis, J. C. and Savin, N. E. (1994), ‘Multiple optima and asymptotic approximations in the partial adjustment model’, *Journal of Econometrics* **62**, 91–128.

Moreira, M. J. (2003*a*), ‘A conditional likelihood ratio test for structural models’, *Econometrica* **71**(4), 1027–1048.

Moreira, M. J. (2003*b*), A general theory of hypothesis testing in the simultaneous equations model, Technical report, Department of Economics, Harvard University, Cambridge, Massachusetts.

Nason, J. M. and Smith, G. W. (2003), Identifying the new Keynesian Phillips curve, Technical report, Univeristy of British Columbia and Queen’s University.

Nelson, C. R. and Startz, R. (1990*a*), ‘The distribution of the instrumental variable estimator and its $t$-ratio when the instrument is a poor one’, *Journal of Business* **63**, 125–140.

Nelson, C. R. and Startz, R. (1990*b*), ‘Some further results on the exact small properties of the instrumental variable estimator’, *Econometrica* **58**, 967–976.

Roberts, J. (1997), ‘Is inflation sticky?’, *Journal of Monetary Economics* **39**, 173–96.

Roberts, J. (1998), Inflation expectations and the transmission of monetary policy, Technical Report 1998-43, Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series, Washington, D.C.

Roberts, J. (2001), How well does the new keynesian sticky-price model fir the data?, Technical Report 2001-13, Board of Governors of the Federal Reserve System, Finance and Economics Discussion Series, Washington, D.C.

Staiger, D. and Stock, J. H. (1997), ‘Instrumental variables regression with weak instruments’, *Econometrica* **65**(3), 557–586.

Stock, J. H. and Wright, J. H. (2000), 'GMM with weak identification', *Econometrica* **68**, 1097–1126.

Stock, J. H., Wright, J. H. and Yogo, M. (2002), 'A survey of weak instruments and weak identification in generalized method of moments', *Journal of Business and Economic Statistics* **20**(4), 518–529.

Wang, J. and Zivot, E. (1998), 'Inference on structural parameters in instrumental variables regression with weak instruments', *Econometrica* **66**(6), 1389–1404.

Woodford, M. (2003), *Interest & Prices: Foundations of a Theory of Monetary Policy*, Princeton University Press, Princeton, New Jersey.

Wright, J. H. (2003), 'Detecting lack of identification in GMM', *Econometric Theory* **19**(2), 322–330.

Zivot, E., Startz, R. and Nelson, C. R. (1998), 'Valid confidence intervals and inference in the presence of weak instruments', *International Economic Review* **39**, 1119–1144.